THE TEXT IS FLY WITHIN THE BOOK ONLY

FUNCTIONS
OF A
COMPLEX VARIABLE

BY

E. J. TOWNSEND, Ph.D.
PROFESSOR OF MATHEMATICS, UNIVERSITY OF ILLINOIS

NEW YORK
HENRY HOLT AND COMPANY
1915

September, 1927

Stanhope Press
F. H. GILSON COMPANY
BOSTON, U.S.A.

PREFACE

The present volume is based on a course of lectures given by the author for a number of years at the University of Illinois. It is intended as an introductory course suitable for first year graduate students and assumes a knowledge of only such fundamental principles of analysis as the student will have had upon completing the usual first course in calculus. Such additional information concerning functions of real variables as is needed in the development of the subject has been introduced as a regular part of the text. Thus a discussion of the general properties of line-integrals, a proof of Green's theorem, etc., have been included. The material chosen deals for the most part with the general properties of functions of a complex variable, and but little is said concerning the properties of some of the more special classes of functions, as for example elliptic functions, etc., since in a first course these subjects can hardly be treated in a satisfactory manner.

The course presupposes no previous knowledge of complex numbers and the order of development is much as that commonly followed in the calculus of real variables. Integration is introduced early, in connection with differentiation. In fact the first statement of the necessary and sufficient condition that a function is holomorphic in a given region is made in terms of an integral. By this order of arrangement, it is possible to establish early in the course the fact that the continuity of the derivative follows from its existence, and consequently the Cauchy-Goursat and allied theorems can be demonstrated without any assumption as to such continuity. Likewise, it can thus be shown that Laplace's differential equation is satisfied without making the usual assumptions as to the existence of the derivatives of second order. The term *holomorphic,* often omitted, has been used as expressing an important property of single-valued functions, reserving the use of the term *analytic* for use in connection with functions derived from a given element by means of analytic continuation. While the Cauchy-Riemann viewpoint is that first introduced, attention is called to the Weierstrass development in the

chapter on series, and in subsequent discussions either definition of an analytic function is used as best suits the purpose in hand.

In Chapter IV much use is made of mapping, thus enabling us to consider in connection with the definition of certain elementary functions some of their more important uses in physics. For the same reason in Chapter V the consideration of linear fractional transformation is especially emphasized and discussed as a kinematic problem. The discussion of series in Chapter VI lays the foundation for the consideration of the fundamental properties of single-valued functions discussed in the following chapter. In the final chapter, it is pointed out how these properties may be extended to the consideration of multiple-valued functions.

The author wishes to express his appreciation of the helpful suggestions which have been given to him by Professor J. L. Markley of the University of Michigan, Professor A. Dresden of the University of Wisconsin, Professor W. A. Hurwitz of Cornell University, and to Dr. Otto Dunkel of the University of Missouri, who have read the proof sheets. He is also under obligations to his colleagues Dr. Denton and Dr. Kempner, who have read the manuscript. Finally, he wishes to express especially his obligations to Dr. George Rutledge, who has rendered him valuable assistance in the preparation of the manuscript.

E. J. TOWNSEND.

University of Illinois
July, 1915

CONTENTS

CHAPTER I

REAL AND COMPLEX NUMBERS

CHAPTER II

FUNDAMENTAL DEFINITIONS CONCERNING FUNCTIONS

CHAPTER III

DIFFERENTIATION AND INTEGRATION

CHAPTER IV

MAPPING, WITH APPLICATIONS TO ELEMENTARY FUNCTIONS

CHAPTER V

LINEAR FRACTIONAL TRANSFORMATIONS

CHAPTER VI

INFINITE SERIES

CHAPTER VII

GENERAL PROPERTIES OF SINGLE-VALUED FUNCTIONS

CHAPTER VIII

MULTIPLE-VALUED FUNCTIONS

FUNCTIONS OF A COMPLEX VARIABLE

CHAPTER I

REAL AND COMPLEX NUMBERS

1. Rational numbers. Some understanding of the nature of a number, the classes into which numbers may be divided, and the general laws governing the fundamental operations with them is essential to the study of the theory of functions. We obtain our first notion of numbers when we undertake to enumerate the individuals composing a group of objects. The process of counting leads, however, only to the positive integers. We arrive at the same result when we assume the existence of unity and a certain mathematical process known as addition. Furthermore, the positive integers obey the following law:

Given any two positive integers a and $b (b > a)$, there exists one and only one positive integer x such that

$$a + x = b.$$

It becomes at once apparent that the positive integers do not completely serve the purpose of analysis when we attempt to solve the above equation for the case where $a = b$. In order to give any interpretation at all to the solution in this case, it is necessary to introduce a new number called zero, defined by the identity

$$a + 0 = a.$$

If a is allowed to be greater than b, it is again necessary to extend the domain of the number-system by the introduction of negative numbers in order to give an interpretation to the solution of the above equation. Even with this extension of the number-system, it is impossible to solve all linear equations. Suppose, for example, it is required to find the value of x from the equation

$$ax = b, \qquad a \neq 0.$$

A number-system that includes only positive and negative integers is inadequate to interpret the result

$$x = \frac{b}{a},$$

whenever b is not an integral multiple of a. A further extension of the number-system now becomes necessary and this extension is gained by the introduction of fractions.

The numbers thus far discussed, that is integers including zero, and fractions, constitute a system of numbers called **rational numbers.*** A characteristic property of such numbers is that they may always be expressed in the form $\frac{b}{a}$, where a and b are integers prime to each other and $a \neq 0$. By the aid of the symbols for the fundamental operations of arithmetic rational numbers can always be expressed by a finite number of digits. It is possible and often convenient to express such numbers by means of an infinite sequence of digits, but it is not necessary to do so. Thus $\frac{1}{3}$ is a rational number, but when expressed in the form of a decimal fraction we have

$$\tfrac{1}{3} = 0.3333 \ldots.$$

2. Irrational numbers. If we undertake to solve equations of a higher degree than the first, the system of rational numbers often proves insufficient. For example, if we have given the equation

$$x^2 - 2 = 0$$

to find the value of x, we have $x = \pm\sqrt{2}$, a result that has no interpretation in the domain of rational numbers. To show that no such interpretation is possible, assume $\frac{a}{b} = \pm\sqrt{2}$, a and b being integers prime to each other. We have then

$$\frac{a^2}{b^2} = 2, \qquad a^2 = 2\,b^2.$$

The number 2 is then a factor of a^2 and as all prime factors appear an even number of times in a perfect square, 2 must appear an even number of times in a^2. Consequently, 2 must also appear as a factor of $2\,b^2$ an even number of times. This, however, is impossible, as it must then appear as a factor of b^2 itself and indeed an even number of times. As 2 cannot be a factor of one member of the identity an even number of times and of the other an odd number of times, the assumption that $\sqrt{2}$ is a rational number is not valid.

* For a more complete discussion of rational numbers the reader is referred to Pierpont, *Theory of Functions of Real Variables,* Vol. I, Chap. I.

We shall see later that it is characteristic of a new class of numbers, called irrational numbers to distinguish them from the numbers discussed in the preceding article, that they do not admit of expression in the form $\frac{a}{b}$.

To see more clearly the nature of irrational numbers, let us consider the totality of rational numbers. Suppose we separate these numbers into two sets such that each number of the first set is greater than every number of the second set. Such a separation of the rational system of numbers is called a **partition.*** We have, for example, a partition if we select any rational number a and put into one set A_1 all those rational numbers that are equal to or greater than a and into a second set A_2 all rational numbers that are less than a. In this case the number a is itself an element of the set A_1.

We may likewise establish a partition by putting into the set A_1 all of those rational numbers greater than a and into A_2 all those equal to or less than a. In this case the number a belongs to set A_2. It will be noticed that by the first partition there is a smallest number in A_1 and by the second partition there is a largest number in A_2. In each case this number is the rational number a itself.

It is possible, however, to establish a partition of the entire system of rational numbers in such a manner that in the one set A_1 there shall be no smallest number and at the same time in the second set A_2, there shall be no largest number. For example, let us consider again $\sqrt{2}$. As we have seen, this number is not a rational number. Put into set A_1 all of those rational numbers whose squares are greater than 2 and into A_2 all rational numbers whose squares are less than 2. The two sets A_1 and A_2 then fulfill the conditions that each number of A_1 is greater than any number of A_2 and there is no smallest number in A_1 and no largest number in A_2; for, no matter how near to 2 the square of a particular rational number may be, there are always other rational numbers whose squares lie between the square of the one selected and 2.

The notion of the partition of the system of rational numbers affords a convenient means of defining irrational numbers. For this purpose suppose the totality of rational numbers to be divided in any manner whatever into two groups A_1, A_2 having the following properties:

* Introduced by Dedekind, *Stetigkeit und irrationale Zahlen*, Braunschweig, 1872.

(1) Each number of the set A_1 shall be greater than any number of the set A_2.

(2) There shall be no smallest number in A_1 and no largest number in A_2.

In the case where a was the smallest rational number in A_1 or the largest one in A_2, it could be said that the partition defined uniquely the rational number a. In the present case, it can no longer be said that the partition defines a rational number; for, every rational number belongs either to set A_1 or set A_2, and since by (2) there can be no smallest number in A_1 and no largest one in A_2, the partition can not define a number in either set. Consequently, the partition may be said to define a new number; we call such a number an **irrational number.** The fundamental operations of arithmetic may be defined for irrational numbers in a manner consistent with the corresponding definitions for rational numbers.*

3. System of real numbers. The rational numbers and the irrational numbers taken together constitute a system of numbers known as **real numbers.** It is this system of numbers that lies at the basis of the calculus of real variables. This system constitutes a closed group with respect to the fundamental operations of arithmetic and obeys certain laws already familiar to the student from his study of algebra. For any numbers a, b, c of this system, we have from the definitions of those fundamental operations

I. *For addition:*

(1) The commutative law: $a + b = b + a$.

(2) The associative law: $a + (b + c) = (a + b) + c$.

II. *For multiplication:*

(1) The commutative law: $ab = ba$.

(2) The associative law: $a(bc) = (ab)\,c$.

(3) The distributive law: $(a + b)\,c = ac + bc$.

(4) Factor law: If $ab = 0$, then $a = 0$ or $b = 0$.

It is customary to introduce subtraction and division as the inverse operations of addition and multiplication. From the definition of these inverse operations and the foregoing fundamental laws follow, as purely formal consequences, all of the rules of operation for real numbers.†

* See Fine, *The Number-System of Algebra*, Art. 29.

† *Ibid.*, Arts. 10, 18.

We assume the existence of a one-to-one correspondence between the totality of real numbers and the points on a straight line; that is to say, we assume that to each real number can be assigned a definite point on the line and conversely to every such point there may be assigned one and only one real number.* This assumption makes possible a geometric interpretation of the results of our discussion and the applications of analysis to geometry.

4. Complex numbers. It will be observed that all real numbers arise from the assumption of a single unit, namely 1. By assuming the additional fundamental unit $\sqrt{-1}$, which we shall represent by i, a very important extension of the number-system thus far discussed can be made. By the use of these two units, 1 and i, we can construct the numbers of the type $a + ib$, where a and b are real numbers. It becomes necessary to extend the number-system so as to include numbers of this type if the solution of the equation

$$ax^2 + bx + c = 0,$$

where $b^2 - 4ac < 0$, is to have any meaning. Such numbers are called **complex numbers** and form the basis of that special branch of the theory of functions to be considered in this volume. It will be seen that since a and b may take all real values, therefore including zero, real numbers are a special case of complex numbers, that is, complex numbers where $b = 0$. In considering the arithmetic of complex numbers, the question arises as to what is to be understood by such terms as "equal to," "greater than," etc., and by the fundamental operations of addition, subtraction, etc. Moreover, it cannot be assumed in advance that the laws of operation with real numbers may be extended without qualification to this broader field. Since real numbers appear as a special case of complex numbers, it is necessary to define these expressions and the fundamental operations in such a manner that the corresponding relations between real numbers shall appear as special cases. These definitions will be considered in the following articles.

Complex numbers involving more than two units have been used by mathematicians. For example, Hamilton, a distinguished English mathematician, introduced higher complex numbers known as

* For references to the mathematical literature where this subject is discussed see: *Encyclopédie des Sciences Mathématiques*, Tome I, Vol. I, pp. 146–147, or Staude's *Analytische Geometrie des Punktes, der geraden Linie, und der Ebene*, p. 422 (10).

quaternions. For this purpose, he made use of the unit 1 and the additional units i, j, k, connected by the following relations:

$$i^2 = j^2 = k^2 = ijk = -1.$$

No use will be made of quaternions or of other higher complex numbers in this volume, and the subject is mentioned merely to illustrate the possibility of further extensions of the number concept.

5. Geometric representation of complex numbers. The assumption which we have made as to the one-to-one correspondence between points on a straight line and the totality of real numbers, makes it possible to give a geometric representation to complex numbers. For this purpose, we introduce a system of rectangular coördinates similar to those used in Cartesian geometry.

To represent the number* $a + ib$, lay off on OX, called the **axis of reals,** the distance a and on OY, called the **axis of imaginaries,** the distance b. Draw through A a line parallel to OY and through B a line parallel to OX. The intersection P of these lines represents the complex number $a + ib$. The numbers a and b may be any real numbers, positive or negative. From these considerations, it follows that there exists a one-to-one correspondence between the points of the plane and the totality of complex numbers. We shall refer to the plane, used in this way, as the **complex plane.** From the relation between the points of the complex plane and the totality of complex numbers, it follows that the complex numbers constitute a continuous system.

By making use of the trigonometric functions, it is possible and frequently convenient to represent complex numbers in another form. From Fig. 1, we have

$$a = \rho \cos \theta, \qquad b = \rho \sin \theta.$$

We may therefore write

$$a + ib = \rho(\cos \theta + i \sin \theta).$$

The distance $OP \equiv \rho$ is called the **modulus** of the complex number, and the angle θ is called the **amplitude** of the complex number.

* The first mathematician to propose a geometric interpretation of the imaginary number $\sqrt{-1}$ was Kühn of Danzig in 1750–1751. The idea was extended by Argand in 1806 to include a representation of complex numbers of the form $a + b\sqrt{-1}$, a representation that was later used by Gauss. The complex plane is frequently referred to as the Argand plane or the Gauss plane.

It will be observed that for any given number $a + ib$ the modulus ρ is a single-valued function of the real numbers a and b, while the amplitude θ is a multiple-valued function of these numbers. The number $a^2 + b^2 = \rho^2$ is frequently referred to as the **norm** of the complex number $a + ib$. The value of θ lying in the interval $-\pi < \theta \leqq \pi$ is called the **chief amplitude.** The amplitude is measured positively in a counter-clockwise direction. The modulus is always to be considered as positive, and hence is often referred to as the absolute value of the complex number. We frequently indicate the modulus or absolute value of any complex number α by placing a vertical line before and after the number, thus $|\alpha|$, read "the absolute value of α."

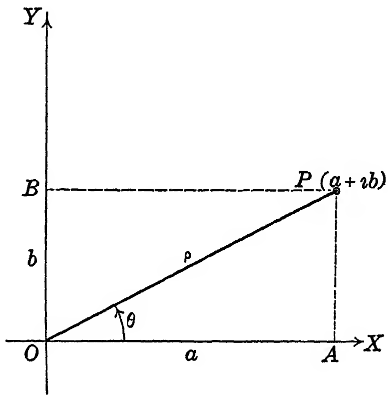

FIG. 1.

Other geometric interpretations of complex numbers are possible. We shall have occasion later to point out, for example, how complex numbers may be represented by points on a sphere by showing that there exists a one-to-one correspondence between the points of the complex plane and those on the surface of a sphere.

From what has already been said, it will be seen that complex numbers are **directed numbers,** that is, numbers that have both magnitude and direction. Consequently, we may when convenient think of the complex number $a + ib$ as represented by the plane vector joining the corresponding point of the complex plane with origin. Such physical magnitudes as force, velocity, acceleration, electric intensity, etc., have direction as well as numerical value and may be represented therefore by complex numbers, provided their directions are confined to a plane. The factor i rotates the given number through an angle $\frac{\pi}{2}$. Thus ia, as we have seen, indicates that a distance a is to be laid off on a line perpendicular to the axis of reals. In the complex number

$$\alpha = \rho(\cos\theta + i\sin\theta),$$

the magnitude of the number is ρ, while the direction in which this

magnitude is measured is determined by the factor in the parenthesis.

6. Comparison of complex numbers. The question very naturally arises as to how two complex numbers may be compared with each other. Given the two numbers

$$\alpha \equiv a + ib, \qquad \beta \equiv c + id.$$

We say that $\alpha = \beta$, when we have the relations

$$a = c, \qquad b = d.$$

Expressed in terms of polar coördinates, equality involves the condition that the two numbers shall have equal moduli and shall have amplitudes that are either equal or differ by some multiple of 2π. It will be observed that the two equal numbers α and β are represented by the same point in the complex plane.

Since the moduli of complex numbers are real, their magnitudes may be compared one with another in the same manner as any other real numbers. Thus of two complex numbers α and β, it is possible to say that the modulus of α is greater than or less than the modulus of β; that is, we may write

$$|\alpha| \gtreqless |\beta|.$$

7. Addition and subtraction of complex numbers. We define the sum of two complex numbers $a + ib$ and $c + id$ as the complex number $(a + c) + i(b + d)$, obtained by adding the real parts and the imaginary parts separately.

It is not to be assumed without proof that the laws of addition enumerated for real numbers in Art. 3 hold for complex numbers. It may be easily shown, however, that such is the case. For this purpose, suppose we have given any three complex numbers

$$\alpha \equiv a + ib, \qquad \beta \equiv c + id, \qquad \gamma \equiv e + if.$$

That the commutative law holds is shown as follows: We have

$$\begin{aligned} \alpha + \beta &= (a + c) + i(b + d) \\ &= (c + a) + i(d + b) \qquad \text{(Art. 3)} \\ &= \beta + \alpha. \end{aligned}$$

To show that the associative law likewise holds we proceed as follows:

$$\begin{aligned} \alpha + (\beta + \gamma) &= [a + (c + e)] + i[b + (d + f)] \\ &= [(a + c) + e] + i[(b + d) + f] \qquad \text{(Art. 3)} \\ &= (\alpha + \beta) + \gamma. \end{aligned}$$

The addition of complex numbers can be easily performed geometrically. In Fig. 2, let $\alpha \equiv a + ib$ and $\beta \equiv c + id$ be represented by the points R and S, respectively. Complete the parallelogram having OR and OS as two sides. The point P then represents the sum $\alpha + \beta$; for, drawing through R a parallel to OX and dropping from P the perpendicular PM, we have from the equality of the two triangles RMP and ONS

$$RM = c, \qquad MP = d,$$

and consequently the coördinates of P are $(a + c,\ b + d)$. Hence, the point P represents the complex number

$$(a + c) + i(b + d) = \alpha + \beta.$$

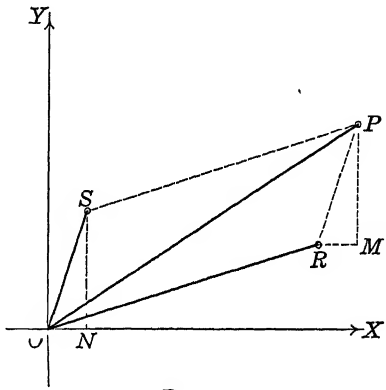

Fig. 2.

The result of geometric addition may be conveniently obtained also as follows: To add β to α, draw from R (Fig. 2) a line parallel to OS, extending in the same direction and equal to it in length. The terminal point P of the line thus drawn represents the number $\alpha + \beta$. To add several numbers in succession, all that is necessary is to draw from the point P representing the sum of the first two numbers a line parallel to the line on which the modulus of the third point is measured, and upon the line thus drawn to lay off from P a segment equal to OP and extending in the same direction. The terminal point of this line represents the sum of the first three numbers. To this result a fourth number may be added in the same way, etc.

An important relation between the absolute values of two complex numbers is suggested by the geometric considerations already introduced. This relation may be stated as follows:

Theorem I. *Given two complex numbers α and β; we have the following relation between their moduli, namely,*

$$\big|\,|\alpha| - |\beta|\,\big| \leqq |\alpha + \beta| \leqq |\alpha| + |\beta|.$$

From elementary geometry, it is known that any side of a triangle is greater than the difference of the other two sides and less than the sum of those sides. Referring now to Fig. 2, we have

$$OR - RP \leqq OP \leqq OR + RP. \tag{1}$$

But we have

$$OR = |\alpha|, \qquad RP = OS = |\beta|, \qquad OP = |\alpha + \beta|. \qquad (2)$$

Hence, substituting these values in (1), we have the result given in the theorem. An equality sign is to be used only when the points representing α and β lie on a straight line passing through the origin.

We may state also the following theorem.

THEOREM II. *For a given complex number* $\alpha \equiv a + ib$, *we have*

$$|a| + |b| \leqq \sqrt{2}\,|a + ib|. \qquad (1)$$

The proof of this theorem, like that for Theorem I, follows at once from geometrical considerations. As the point α moves around the circle (Fig. 3) $|a|$ and $|b|$ change values. Let us find the maximum value of $|a| + |b|$, subject to the condition that

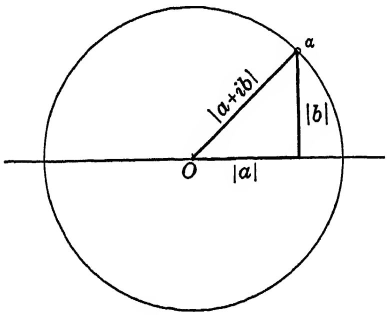

FIG. 3.

$$|a|^2 + |b|^2 = |a + ib|^2 = r^2, \qquad (2)$$

where r is the radius of the given circle. By the ordinary methods for computing a maximum, we find that the function in question has a maximum if

$$|a| = |b| = \frac{|a + ib|}{\sqrt{2}}.$$

Consequently, the maximum value of $|a| + |b|$ is $\sqrt{2}\,|a + ib|$, from which the relation (1) must hold for the various values of α upon the circle.

We define **the difference** $\alpha - \beta$ of two complex numbers, $\alpha \equiv a + ib$, $\beta \equiv c + id$, as the complex number $(a - c) + i(b - d)$. Here, as in addition, the laws of operation with real numbers hold for complex numbers and follow as a consequence of the laws of addition and the definition of subtraction. The reader can easily verify this statement.

Let α and β be represented by R and S, respectively (Fig. 4). To subtract β from α geometrically, construct the parallelogram having OR as a diagonal and OS as one side. The point P represents $\alpha - \beta$, since the sum of β and the complex number represented by this point is α. The same result is obtained by drawing the line RP parallel to SO and equal to it in length.

8. Multiplication of complex numbers. We may define the product of two complex numbers α, β by the relation

$$\alpha\beta = (a + ib)\,(a' + ib')$$
$$= (aa' - bb') + i(ab' + ba').$$

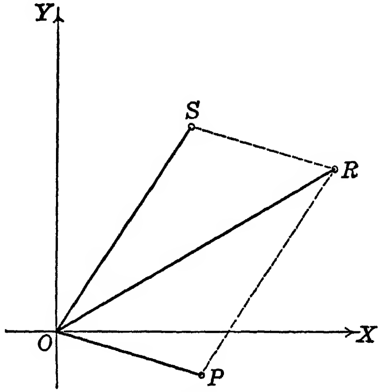

FIG. 4.

If the given numbers are written in the form

$$\alpha = \rho_1(\cos\theta_1 + i\sin\theta_1),$$
$$\beta = \rho_2(\cos\theta_2 + i\sin\theta_2),$$

then the product $\alpha\beta$ becomes

$$\alpha\beta = \rho_1(\cos\theta_1 + i\sin\theta_1)\cdot\rho_2(\cos\theta_2 + i\sin\theta_2)$$
$$= \rho_1\rho_2[(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2)]$$
$$= \rho_1\rho_2[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)].$$

This relation gives us the rule for multiplication, which may be stated in words as follows:

The product of two complex numbers is a complex number whose modulus is the product of the moduli of the two given numbers and whose amplitude is the sum of the given amplitudes.

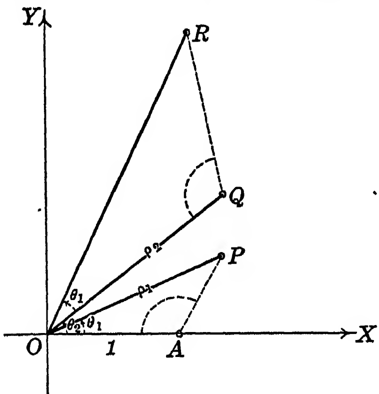

FIG. 5.

The associative, commutative, distributive, and factor laws for multiplication hold for complex numbers as for real numbers. The commutative law for example can be established as follows:

Given as before α, β, we have

$$\alpha\beta = \rho_1\rho_2[\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)]$$
$$= \rho_2\rho_1[\cos(\theta_2+\theta_1) + i\sin(\theta_2+\theta_1)]$$
$$= \beta\alpha. \qquad \text{(Art. 3)}$$

In a similar manner the associative, distributive, and factor laws may be established.

From the definition of a product, we are able easily to give a geometric interpretation of multiplication. Represent the two complex numbers α and β by the points P and Q, respectively. Thus far we have been able to carry out the geometric operations introduced without reference to the magnitude of the unit. We must now

make use of a unit length. Lay off on OX the distance OA, which we take as the unit of length. Connect the point P with A. On OQ as a base construct the triangle OQR similar to the triangle OAP. We have then from the figure

$$OR : OP :: OQ : 1;$$

that is,

$$OR : \rho_1 :: \rho_2 : 1,$$

or

$$\rho_1\rho_2 = OR.$$

Furthermore, we have by construction

$$\angle QOR = \angle AOP;$$

hence

$$\angle AOR = \theta_1 + \theta_2.$$

Consequently, from the definition of multiplication it follows that the point R represents geometrically the required product; for, it has the modulus $\rho_1\rho_2$ and the amplitude $\theta_1 + \theta_2$.

The rule of multiplication may be extended to the product of any finite number of complex numbers. Suppose, for example, we have

$$\begin{aligned} \alpha_1 &= \rho_1(\cos\theta_1 + i\sin\theta_1), \\ \alpha_2 &= \rho_2(\cos\theta_2 + i\sin\theta_2), \\ &\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \alpha_n &= \rho_n(\cos\theta_n + i\sin\theta_n). \end{aligned}$$

We obtain as the continued product of these numbers

$$\alpha_1\alpha_2 \ldots \alpha_n = \rho_1\rho_2 \ldots \rho_n[\cos(\theta_1 + \cdots + \theta_n) + i\sin(\theta_1 + \cdots + \theta_n)]$$

When $\rho_1 = \rho_2 = \cdots = \rho_n = 1$, we have an important theorem, known as **De Moivre's theorem,** namely:

$$\begin{aligned} &(\cos\theta_1 + i\sin\theta_1)\cdot(\cos\theta_2 + i\sin\theta_2) \ldots (\cos\theta_n + i\sin\theta_n) \\ &= \cos(\theta_1 + \theta_2 + \cdots + \theta_n) + i\sin(\theta_1 + \theta_2 + \cdots \theta_n). \end{aligned}$$

If we also put $\theta_1 = \theta_2 = \cdots = \theta_n = \theta$, we have the form of the theorem most frequently used, namely:

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta.$$

This theorem gives us a method of raising any complex number to an integral power; for, we have

$$[\rho(\cos\theta + i\sin\theta)]^n = \rho^n(\cos n\theta + i\sin n\theta).$$

The power of a complex number may be found geometrically as follows: Let P_1 represent the number

$$\alpha = \rho(\cos\theta + i\sin\theta).$$

On OP_1 construct the triangle OP_1P_2 similar to OAP_1; then P_2 represents the second power of the given complex number. On OP_2 as a base construct another triangle OP_2P_3 similar to OAP_1 or to OP_1P_2. The point P_3 of this triangle represents the cube of the given number. Continuing in this way, we may raise the given number to any required integral power.

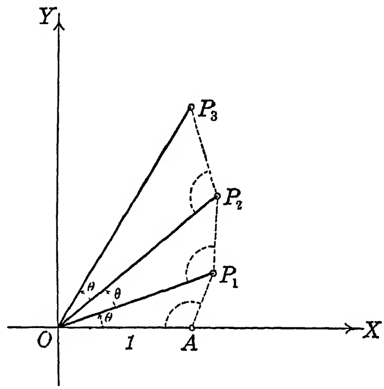

FIG. 6.

It may be shown that De Moivre's theorem holds likewise for negative and fractional powers.* As an illustration, let us consider the n^{th} roots of a complex number α. We have then

$$\alpha^{\frac{1}{n}} = \rho^{\frac{1}{n}}\left(\cos\frac{\theta}{n} + i\sin\frac{\theta}{n}\right),$$

where $\rho^{\frac{1}{n}}$ is always to be considered positive and real. We obtain the remaining n^{th} roots of α by replacing θ by $\theta + 2k\pi$, $k = 1, 2, 3, \ldots, n-1$. Thus the n roots are

$$\rho^{\frac{1}{n}}\left(\cos\frac{\theta}{n} + i\sin\frac{\theta}{n}\right),$$

$$\rho^{\frac{1}{n}}\left(\cos\frac{\theta+2\pi}{n} + i\sin\frac{\theta+2\pi}{n}\right),$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$\rho^{\frac{1}{n}}\left(\cos\frac{\theta+2(n-1)\pi}{n} + i\sin\frac{\theta+2(n-1)\pi}{n}\right).$$

That each of these numbers is an n^{th} root of α is seen at once from the fact that its n^{th} power gives the number

$$\alpha = \rho(\cos\theta + i\sin\theta).$$

* See Hobson, *Plane Trigonometry*, 2d Ed. Arts. 181 and 186.

If θ be the chief amplitude of α, that is if $-\pi < \theta \leqq \pi$, then

$$\rho^{\frac{1}{n}}\left(\cos\frac{\theta}{n} + i\sin\frac{\theta}{n}\right)$$

is called the chief or **principal value of the root.**

For example, consider the positive real number $\alpha = a$. The two square roots of α are

$$\sqrt{a}(\cos 0 + i\sin 0), \qquad \sqrt{a}(\cos\pi + i\sin\pi).$$

The principal value of $\alpha^{\frac{1}{2}}$ is

$$\sqrt{a}(\cos 0 + i\sin 0) = \sqrt{a};$$

for, in this case

$$\theta = n \cdot \operatorname{amp}\alpha^{\frac{1}{n}} = 2\cdot 0 = 0$$

falls in the interval $-\pi < \theta \leqq \pi$.

If we consider the number $\alpha = -a$, we have as the two square roots

$$\sqrt{a}\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right), \qquad \sqrt{a}\left(\cos\frac{3\pi}{2} + i\sin\frac{3\pi}{2}\right),$$

and the principal value of $\alpha^{\frac{1}{2}}$ is

$$\sqrt{a}\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right) = i\sqrt{a}.$$

As a further illustration of the use of De Moivre's theorem, let us consider the n roots of unity. Here $\theta = 0$, and we have as the roots

$$\cos 0 + i\sin 0,$$

$$\cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n},$$

$$\cos\frac{4\pi}{n} + i\sin\frac{4\pi}{n},$$

$$\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot$$

$$\cos\frac{2(n-1)\pi}{n} + i\sin\frac{2(n-1)\pi}{n}.$$

If we denote the second of these roots by ω, then the n roots may be written

$$1,\ \omega,\ \omega^2,\ \ldots,\ \omega^{n-1},$$

Since we have

$$\cos\frac{\theta+2\,k\pi}{n}+i\sin\frac{\theta+2\,k\pi}{n}=\left(\cos\frac{\theta}{n}+i\sin\frac{\theta}{n}\right)\left(\cos\frac{2\,k\pi}{n}+i\sin\frac{2\,k\pi}{n}\right), \quad (1)$$

we may write the n roots of any complex number α in the form

$$\alpha^{\frac{1}{n}},\ \omega\alpha^{\frac{1}{n}},\ \omega^2\alpha^{\frac{1}{n}},\ \ldots,\ \omega^{n-1}\alpha^{\frac{1}{n}}, \quad (2)$$

where $\alpha^{\frac{1}{n}}$ denotes one of the n roots of α, for example the principal value of the root; that is to say, the n roots of any complex number are the products of some one value of the root into the n roots of unity.

In certain cases the roots of a complex number may be determined graphically. It is not possible, however, to do this in all cases; for, it is not always possible to make the construction by means of the ruler and compasses. Let us consider the fourth roots of the number α, represented by the point P (Fig. 7). Denote the modulus of α by ρ and its chief amplitude by θ. The principal value of the root is then given by

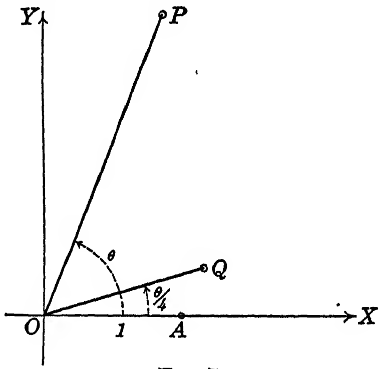

Fig. 7.

$$\alpha^{\frac{1}{4}}=\rho^{\frac{1}{4}}\left(\cos\frac{\theta}{4}+i\sin\frac{\theta}{4}\right).$$

To determine this root, it is necessary to construct an angle $\frac{\theta}{4}$ and to lay off a distance equal to $\rho^{\frac{1}{4}}$. We can construct the angle $\frac{\theta}{4}$ by dividing twice in succession the angle θ by the methods of elementary geometry. We can find the line segment that represents $\rho^{\frac{1}{4}}$ by constructing the mean proportional between OP and 1 and then constructing the mean proportional between that result and 1. In this manner the point Q is determined representing the principal value of the fourth root of α. That Q does represent a fourth root of α, may be shown by constructing, as indicated in Fig. 6, the fourth

power of the number represented by Q. There are three other fourth roots of α. From (2) they are seen to be

$$\alpha^{\frac{1}{4}}\omega, \quad \alpha^{\frac{1}{4}}\omega^2, \quad \alpha^{\frac{1}{4}}\omega^3,$$

where ω denotes a fourth root of unity. To multiply by ω, ω^2, or ω^3 is to rotate the line OQ in a counter-clockwise direction through an angle of 90°, 180°, 270°, respectively, about O as a center. To find geometrically the four points representing the fourth roots of α is equivalent to constructing a regular inscribed polygon of four sides in a circle having OQ as a radius and Q as one of the vertices. Each vertex of this polygon is a fourth root of α, as may be verified by constructing its fourth power.

To determine graphically all of the n^{th} roots of any complex number involves the division of the chief amplitude into n equal parts, the laying off of a distance equal to the n^{th} root of the given modulus, and the inscribing of a regular polygon in a circle. For the special case of the n^{th} roots of unity, the problem reduces to the construction of a regular polygon inscribed in a circle of unit radius; for, the modulus is 1 and in this case the chief amplitude is zero. As has already been pointed out this construction is not always possible by means of a ruler and compasses. The construction is however possible if and only if we have * $n = 2^l p_1 p_2 \ldots$, where l is an integer and $p_1, p_2 \ldots$ are distinct prime numbers of the form $2^{2^t} + 1$. For example, it is possible if

$$n = 3, 4, 5, 6, 8, 10, 12, 15, 16, 17, 20, 24, \ldots$$

and impossible if

$$n = 7, 9, 11, 13, 14, 18, 19, 21, 22, 23, 25, \ldots.$$

9. Division of complex numbers. Given two complex numbers

$$\alpha \equiv a + ib, \qquad \beta \equiv a' + ib'.$$

Since division is the inverse operation of multiplication, we must so define the quotient of α by β that the result multiplied by β gives α. In accordance with this relation, we define the quotient of α divided by β by means of the identity

$$\frac{\alpha}{\beta} = \frac{a + ib}{a' + ib'} = \frac{aa' + bb'}{a'^2 + b'^2} + i\frac{a'b - ab'}{a'^2 + b'^2}.$$

* See *Monographs on Modern Mathematics*, edited by J. W. A. Young, p. 379.

Writing α, β in the form

$$\alpha = \rho_1(\cos\theta_1 + i\sin\theta_1), \qquad \beta = \rho_2(\cos\theta_2 + i\sin\theta_2),$$

the quotient of α by β may be written as follows:

$$\frac{\alpha}{\beta} = \frac{\rho_1}{\rho_2}[\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)].$$

This form of the definition may be expressed in words as follows:

The quotient of one complex number by another is a complex number whose amplitude is the amplitude of the dividend minus that of the divisor, and whose modulus is the modulus of the dividend divided by that of the divisor.

We have already pointed out in another connection that the amplitude of a complex number is multiple-valued. This, however, does not affect the quotient; for, an increase of the amplitude of the dividend or the divisor by a multiple of 2π increases or decreases likewise the amplitude of the quotient by the same multiple of 2π and hence the result remains unchanged.

From the definition of division, we have for the reciprocal of a complex number α

$$\begin{aligned}\frac{1}{\alpha} &= \frac{(\cos 0 + i\sin 0)}{\rho(\cos\theta + i\sin\theta)}\\ &= \frac{1}{\rho}[\cos(-\theta) + i\sin(-\theta)]\\ &= \frac{1}{\rho}(\cos\theta - i\sin\theta).\end{aligned}$$

We may perform geometrically the operation of division as follows: Let P, Q, represent the two complex numbers

$$\alpha = \rho_1(\cos\theta_1 + i\sin\theta_1)$$

and

$$\beta = \rho_2(\cos\theta_2 + i\sin\theta_2),$$

respectively. Draw the line OM (Fig. 8) making the angle $-\theta_2$ with OP. Construct on OP the triangle ORP similar to OQA, when $OA = 1$. The point R represents the quotient; for, it has the amplitude $\theta_1 - \theta_2$ and its modulus is $\frac{\rho_1}{\rho_2}$, since we have from the two similar triangles, ORP and OAQ,

$$OR : 1 :: \rho_1 : \rho_2,$$

and, hence,

$$\frac{\rho_1}{\rho_2} = OR.$$

It will be observed that $\dfrac{\alpha}{\beta}$ has no significance when $\rho_2 = 0$; for, division by zero is meaningless. The general laws of division for real numbers hold for complex numbers and are a consequence of the definition of division and the laws of operation governing multiplication.

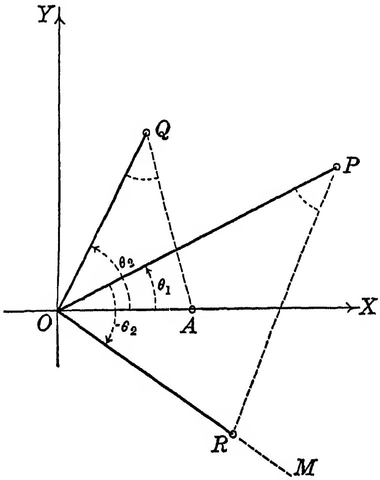

FIG. 8.

We have now defined the fundamental operations of arithmetic with reference to complex numbers. Moreover, we have seen that the general laws of operation in the arithmetic of real numbers may be extended without modification to complex numbers. We are now in a position to introduce the complex variable and functions of it. Certain fundamental notions concerning the functions to be considered will be discussed in the next chapter.

EXERCISES

1. Express $\dfrac{\alpha\beta}{\gamma} + \gamma^2$ in the form $A + iB$, where α, β, γ are given complex numbers.

2. Perform graphically the operations indicated by $(\alpha\beta - \gamma) \div \gamma$, $(\alpha\beta \div \gamma) - 1$, where $\alpha \equiv 2 + i\,3$, $\beta \equiv 1 + i\,2$, $\gamma \equiv 3 - i\,2$.

3. Represent graphically the square roots of $3 + i\,2$, $-2 + i\,3$, $\frac{1}{3} + i\,\frac{1}{4}$.

4. Represent geometrically the four values of $(1 + \sqrt{-1})^{\frac{1}{4}}$.

5. Locate the points representing the sixth roots of 1 and of -1.

6. Give an illustration of two complex numbers the sum of whose moduli is equal to the modulus of their sum; also two complex numbers the sum of whose moduli is greater than the modulus of their sum.

7. Under what conditions do the relations

$$|a + b| = |a| + |b|,$$
$$|a + b| = |a| - |b|,$$

hold when a, b are real numbers? Under what conditions do they hold when a, b are replaced by the complex numbers α, β?

8. Prove the associative and distributive laws for multiplication of complex numbers.

9. If α and β are complex numbers, where $\alpha \neq 0$, $\beta \neq 0$, prove that $\alpha\beta \neq 0$.

10. Interpret geometrically $(\alpha\beta)^m = \alpha^m\beta^m$; $\alpha^m\alpha^n = \alpha^{m+n}$.

11. A boat is being rowed from the west to the east bank of a stream at the rate of three miles per hour. At the same time it is being carried north by the current at the rate of two miles per hour. Represent the velocity by a point in the complex plane. What represents the speed?

12. A fly-wheel two feet in diameter is revolving counter-clockwise at the rate of 180 revolutions per minute. Express as a complex number of the form $a + ib$ the velocity of a point on the rim of the wheel where the radius vector through that point makes an angle ϕ with a fixed initial position.

CHAPTER II

FUNDAMENTAL DEFINITIONS CONCERNING FUNCTIONS

10. Constants, variables. We shall make the same distinctions between constants and variables as in the realm of real variables. If a complex number assumes but a single value in any discussion, it is called a **constant**. The numbers thus far considered serve as illustrations. If, on the other hand, a number is allowed in any discussion to assume various complex values, it is called a **variable**. A complex variable z may be written in the form $x + iy$, where x and y are real variables.

We shall speak of a connected portion of the complex plane as a **region** or domain. Any point z_0 is said to be an **inner point** of a region if it can be made the center of a circle of radius different from zero such that all points within this circle are points of the region. If the circle can be taken so small that it contains no points of the region, then z_0 lies exterior to the region. If the circle contains both points of the region and points exterior to it, however small the radius of the circle be taken, then z_0 is called a **boundary point** of the region. If the boundary is included in the region, then it is spoken of as a **closed region,** otherwise it is called an **open region.** Unless the contrary is stated, the term region will be used to designate an open region, that is, a connected portion of the complex plane consisting only of inner points. A given region may be finite or it may be infinite. The inner points of an infinite region may all lie exterior to a given curve in which case the curve is the boundary of the region if its points are boundary points. By a **neighborhood** of a given point z_0, we shall understand a small region about z_0 having z_0 as an inner point. For most purposes it will be found convenient to choose a neighborhood for which

$$|z - z_0| < \rho,$$

that is a circle of radius ρ about the point z_0. In referring to the points of a neighborhood of z_0 exclusive of the point z_0 itself, we shall speak of the region as a **deleted neighborhood** of z_0.

11. Definition and classification of functions. The definition of a function has undergone radical changes since it was first introduced in connection with real variables. Leibnitz, for example, associated the term with any expression standing for certain lengths connected with curves, such as tangents, radii of curvature, normals, etc. Euler, who wrote the first treatise on the theory of functions, defined a function as an analytic expression in which one or more variables appear. We must set aside such a definition because there are numerous illustrations of related and interdependent variables, both in pure mathematics and in theoretical physics, where as yet no analytic expression has been found. These early definitions of a function also assumed that a continuous function can always be represented geometrically by a continuous curve having a definite tangent at each point. This condition involves the requirement that every continuous function shall have a definite derivative for each value of the variable. Subsequent researches show that this is an erroneous assumption, and that there exist functions defined by analytic expressions that are continuous throughout an interval and that do not possess a derivative at any point of that interval. The study of Fourier's theory of heat led Dirichlet in 1837 to formulate the following definition of a function of a real variable, which is the one commonly accepted by mathematicians at the present time: *

If for each value of a variable x, there is determined a definite value or set of values of another variable y, then y is called a function of x for those values of x.

This definition does not necessitate the existence of any analytic relation between y and x. It is to be observed that it is not necessary that x should have every value in an interval; it may take, for example, only a set of values. What is essential in the definition is that for every value that x does take, there is thereby determined a definite value or definite values of y. An important step in the evolution of the idea of a function was made when Cauchy gave to the variable complex values, and extended the notion of a definite integral by letting the variable pass from the one limit of integration to the other through a succession of complex values along arbitrary paths. The work of Cauchy and the subsequent work of Riemann and Weierstrass laid the foundation for the development of the general theory of functions.

* See *Encyclopédie des Sciences Math.,* Tome II, Vol. I, Fasc. 1, p. 13.

We shall understand the complex variable w to be a function of the complex variable z in a given open or closed region S if for each value of z in this region w has a definite value or set of values. Here, as in the case of functions of a real variable, the function may be defined also with respect to a set of values rather than for all values of z of a given region. Unless otherwise stated, however, it will be understood that z takes all values of a given region. Then, if w is a function of z, we may write

$$w = u(x, y) + i\,v(x, y) = f(z),$$

where u, v are real functions of the two real variables x and y. If w has but one value for each value of z, w is said to be **single-valued;** if it takes two or more values for some or all of the values of z, then w is called a **multiple-valued function** of z.

A possible criticism of Dirichlet's definition is that it is not sufficiently restrictive. Without introducing some additional properties, such as continuity, differentiability, etc., it is impossible upon the basis of this definition of a function to build a theory of analysis permitting the operations and developments similar to those of the calculus of real variables. In later articles we shall discuss certain characteristic properties that the functions to be considered in this volume must possess. Before doing so we shall define certain general classes of functions and discuss some of the fundamental conceptions that are of importance in the consideration of their properties.

One of the important classifications of functions is their division into rational functions and irrational functions. By a **rational integral function** or polynomial is understood a function of the form

$$\alpha_n z^n + \alpha_{n-1} z^{n-1} + \cdots + \alpha_0,$$

where $\alpha_0, \alpha_1, \ldots, \alpha_n$ are constants and n is a positive integer. If $\alpha_n \neq 0$, the function is said to be of the **n^{th} degree.**

A **rational fractional function** is the quotient of two rational integral functions having no common factor and hence it is of the form

$$\frac{\alpha_n z^n + \alpha_{n-1} z^{n-1} + \cdots + \alpha_0}{\beta_m z^m + \beta_{m-1} z^{m-1} + \cdots + \beta_0},$$

where m and n may be equal or different positive integers. If $\alpha_n \neq 0$, $\beta_m \neq 0$, and $m = n$, then this common value is called the **degree** of the function; if $m \neq n$, then the larger of the two is called the degree.

All functions which are not rational are classified as **irrational functions.**

Another important classification of functions is into algebraic and transcendental. We say that w is an **algebraic function** of z when w and z are related by an irreducible equation of the form

$$f_0(z)\,w^n + f_1(z)\,w^{n-1} + f_2(z)\,w^{n-2} + \cdots + f_n(z) = 0,$$

where $f_0(z), f_1(z), f_2(z), \ldots, f_n(z)$ are rational integral functions of z. We say that this equation is irreducible if it is not possible to write the left-hand member as the product of two polynomials, neither of which is a constant. It will be seen that all rational functions, for example, are algebraic functions.

All functions that are not algebraic are called **transcendental functions.** Such functions include the trigonometric, exponential, and logarithmic functions.

12. Limits. From the study of elementary mathematics, and particularly from the study of the calculus, the student is familiar with the general notion and properties of limits. We shall recall some of the fundamental properties by way of emphasis and extend the notion of a limit to the realm of complex numbers.

If we have given, for example, the sequence of numbers

$$2,\ 1\tfrac{1}{2},\ 1\tfrac{1}{3},\ \ldots,\ 1\tfrac{1}{n},\ \ldots,$$

it is at once seen that by taking n sufficiently large the terms of the sequence can be made to ultimately differ from unity by as little as we please. We express this fact by saying that the sequence has the limit 1. Likewise, the sequences

$$1,\ \tfrac{1}{4},\ \tfrac{1}{6},\ \ldots,\ \tfrac{1}{2n},\ \ldots$$

$$1,\ \tfrac{1}{4},\ \tfrac{1}{9},\ \ldots,\ \tfrac{1}{n^2},\ \ldots$$

have the limit zero. If we may assign at pleasure to a number values which are numerically as small as we may choose, then the number is said to be **arbitrarily small.** We shall usually denote such a number by ϵ. We may now define the limit of a sequence more exactly as follows:

Suppose we have given an infinite sequence of real numbers

$$\{a_n\} \equiv a_1,\ a_2,\ a_3,\ \ldots,\ a_n,\ \ldots$$

If there exists a definite number a, and, corresponding to an arbitrarily small positive number ϵ, a positive integer m such that for all values of $n > m$, we have

$$|a_n - a| < \epsilon,$$

then a is called the **limit of the sequence,** and we write

$$\underset{n=\infty}{L}\, a_n = a.$$

If we have given an infinite sequence of complex numbers, the moduli of these complex numbers form a sequence of real numbers, and so do the moduli of the differences between these complex numbers and any complex constant. We say that a sequence of complex numbers

$$\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n, \ldots$$

has the limit α or converges to the limit α, if the moduli of the differences between these complex numbers and α form a sequence having the limit zero; that is, if corresponding to an arbitrarily small positive number ϵ, it is possible to find a positive integer m such that we have

$$|\alpha_n - \alpha| < \epsilon, \qquad n > m. \tag{1}$$

We then write

$$\underset{n=\infty}{L}\, \alpha_n = \alpha.$$

Since the relation given by (1) holds for all integral values of $n > m$, it likewise holds for a particular set of values of $n > m$, say for the even values of $n > m$. In other words, any subsequence selected from the given sequence will have the same limiting value as the given sequence.

The foregoing definition of the limit of a sequence may be expressed in terms of a and b, where

$$\alpha = a + ib, \qquad \alpha_n = a_n + ib_n;$$

for, we have the following theorem.

THEOREM I. *The necessary and sufficient condition that the sequence of complex numbers*

$$\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n, \ldots$$

converges to a limit $\alpha = a + ib$ *is that*

$$\underset{n=\infty}{L}\, a_n = a; \qquad \underset{n=\infty}{L}\, b_n = b. \tag{2}$$

We have

$$\alpha_n - \alpha = a_n + ib_n - a - ib$$
$$= (a_n - a) + i(b_n - b),$$

whence

$$|\alpha_n - \alpha| \leqq |a_n - a| + |b_n - b|, \tag{3}$$

and

$$|a_n - a| \leqq |\alpha_n - \alpha|, \qquad |b_n - b| \leqq |\alpha_n - \alpha|. \tag{4}$$

The condition stated in the theorem is necessary; for, if the given sequence has the limit α, we may write

$$|\alpha_n - \alpha| < \epsilon$$

for n sufficiently great. Hence, from (4) we have also

$$|a_n - a| < \epsilon, \qquad |b_n - b| < \epsilon;$$

that is,

$$\underset{n=\infty}{L}\, a_n = a, \qquad \underset{n=\infty}{L}\, b_n = b.$$

The given condition is also sufficient; for, if the two limits (2) exist, we thus have for n sufficiently great.

$$|a_n - a| < \epsilon, \qquad |b_n - b| < \epsilon,$$

and hence from (3) it follows that

$$|\alpha_n - \alpha| < 2\,\epsilon.$$

Therefore, the given sequence has the limit α as the theorem requires.

Suppose the variable z takes a set of values dense at α, that is, a set of values such that in every neighborhood of α, however small, there are an infinite number of points representing values of z. We express this relation between z and α by saying that α is a **limiting point** of the variable z. Under these conditions the variable z may be said to approach its limiting value α; that is, it may so vary that $|z - \alpha|$ decreases indefinitely. When z varies in this manner, we write $z \doteq \alpha$, which is to be read "as z approaches α." In the discussions to follow the given set of values of z will usually include every value in the neighborhood of α, and unless otherwise stated this will be understood to be the case.

As the variable z takes different values, any single-valued function of z, say $f(z)$, has by definition a definite value for each value z is allowed to take. The values of $f(z)$, corresponding to the values of z in a suitably small deleted neighborhood of a limiting point α, may likewise differ from some number A by amounts whose numerical values are less than an arbitrarily small positive number. We then

speak of the number A as the limiting value of the function $f(z)$ corresponding to the limit α of z. We may also say that $f(z)$ approaches A as z approaches α; for, under the conditions just stated, no matter how z may approach α, $f(z)$ will at the same time approach A. We may now formulate the definition of the **limit of a function** as follows:

If α is a limiting point of z, and if corresponding to an arbitrarily small positive number ϵ there exists a positive number δ such that for all values of z entering into the discussion for which $|z - \alpha| < \delta$, with the possible exception of $z = \alpha$, we have

$$|f(z) - A| < \epsilon, \tag{5}$$

then $f(z)$ is said to have the limiting value A corresponding to the limit α of z. We indicate the existence and the value of this limit by writing

$$\underset{z \doteq \alpha}{L} f(z) = A. \tag{6}$$

The limit of a function does not depend upon the value of the function for the limiting value of the variable, but only upon the values that the function takes in the deleted neighborhood of such a point. Frequently the value of the function at the point is quite different from its limiting value. So far as the mere existence of the limit is concerned, it is not essential that the function be defined for the limiting value of the variable. The general laws of operation with limits as developed in the algebra and used in the calculus of real variables hold equally well for complex variables, as they are developed without reference to any particular domain of numbers.

As we have already seen, the existence of the limit of a function involves the condition that the same limiting value of $f(z)$ is obtained whatever be the set of values through which z is permitted to approach the critical value α. As z may be written in the form $x + iy$, where x and y are independent real variables and α is a number of the form $a + ib$, it will be seen at once that the limit given in (6) is related to the double simultaneous limit $\underset{\substack{x \doteq a \\ y \doteq b}}{L} F(x, y)$, discussed in connection with functions of two real variables,* the existence of which requires that the same limiting value be obtained by all possible methods of approach of the variable point (x, y) to limiting position (a, b).

* See Townsend and Goodenough, *First Course in Calculus*, Arts. 101 and 102.

A necessary and sufficient condition for the existence of the limit (6) may be stated as follows:

THEOREM II. *Given $z = x + iy$, $\alpha = a + ib$, $\beta = A + iB$; the necessary and sufficient conditions that $f(z) \equiv u + iv$ approaches β as z approaches α are that*

$$\underset{\substack{x \doteq a \\ y = b}}{L} u(x, y) = A, \qquad \underset{\substack{x \doteq a \\ y = b}}{L} v(x, y) = B. \tag{7}$$

To prove that the conditions stated in the theorem are necessary, we have given

$$\underset{z \doteq \alpha}{L} f(z) = \beta, \tag{8}$$

to show that the two limits (7) exist. From (8), we have

$$|f(z) - \beta| < \epsilon, \qquad |z - \alpha| < \delta.$$

This relation may be written

$$|u + iv - A - iB| < \epsilon, \tag{9}$$

for values of (x, y) that lie within the circle of radius δ about the point (a, b); that is, for $\sqrt{(x-a)^2 + (y-b)^2} < \delta$. By aid of Theorem II of Art. 7, we now have from (9)

$$|u - A| + |v - B| \leqq \sqrt{2}\,|(u - A) + i(v - B)| < \epsilon\sqrt{2}.$$

Hence, it follows that

$$|u - A| < \epsilon\sqrt{2}, \quad |v - B| < \epsilon\sqrt{2}, \quad \sqrt{(x-a)^2 + (y-b)^2} < \delta.$$

Expressed in the form of limits, this result is

$$\underset{\substack{x \doteq a \\ y \doteq b}}{L} u(x, y) = A, \qquad \underset{\substack{x \doteq a \\ y \doteq b}}{L} v(x, y) = B.$$

We may show as follows that the given conditions are also sufficient. We have given the two limits (7) to show the existence of the limit (8). Expressed as inequalities, the limits in (7) give

$$|u(x, y) - A| < \epsilon, \qquad \sqrt{(x-a)^2 + (y-b)^2} < \delta_1, \tag{10}$$

$$|v(x, y) - B| < \epsilon, \qquad \sqrt{(x-a)^2 + (y-b)^2} < \delta_2. \tag{11}$$

From Theorem I of Art. 7, we have

$$|(u - A) + i(v - B)| \leqq |u - A| + |v - B|. \tag{12}$$

By use of (10) and (11) this relation becomes

$$|u + iv - A - iB| < 2\epsilon, \qquad \sqrt{(x-a)^2 + (y-b)^2} < \delta', \quad (13)$$

where δ' is the smaller of the two numbers δ_1, δ_2.

Hence, the relation (13) may be written

$$|f(z) - \beta| < 2\epsilon, \qquad |z - \alpha| < \delta'.$$

Expressing this result in terms of a limit, we have

$$\underset{z=\alpha}{L} f(z) = \beta,$$

as the theorem requires.

If for a set of real numbers there exists a definite number A such that the numbers of the set never exceed A but are dense at A, then A is called the **upper limit** of the set. For example, the set of elements constituting the sequence

$$1 - \tfrac{1}{2},\ 1 - \tfrac{1}{3},\ 1 - \tfrac{1}{4},\ \ldots,\ 1 - \tfrac{1}{n},\ \ldots$$

has the upper limit 1. In this particular case the number 1 is not an element of the set, although the elements approach the value 1. It is possible that the upper limit of a set shall be itself an element of the set. When this is the case, we call the upper limit of the set the **maximum value** of the set. Thus, in the sequence just given 1 is the upper limit of the set but not the maximum value of the set.

Likewise, if none of the elements of a set of real values are smaller than a definite number A and the set is dense at A, then the number A is called the **lower limit** of the set. If the lower limit is at the same time an element of the set, it is called the **minimum value** of the set. For example, elements of the sequence

$$3,\ 2\tfrac{1}{2},\ 2\tfrac{1}{3},\ 2\tfrac{1}{4},\ \ldots,\ 2\tfrac{1}{n},\ \ldots$$

form a set having the lower limit 2. It does not, however, have a minimum value as 2 is not an element of the set. In other words there is no smallest number in the set.

While the terms, upper limit, lower limit, maximum, minimum, are defined with reference to a set of real values, they may be extended without modification to the absolute values of a function $f(z)$ of a complex variable.

If the elements of a given sequence all lie in a finite interval, the sequence is often spoken of as a **bounded sequence.** Every bounded

sequence of increasing real numbers has a limit.* If the sequence is not bounded, but its elements increase without limit, we say that the sequence becomes infinite and indicate that fact frequently by writing the limit equal to $+\infty$. The sign of equality in this connection is not to be confused with the ordinary use of that symbol and should be understood as a brief and convenient way of writing "becomes infinite" or "increases without limit." If the elements of a sequence decrease without limit, we place the limit equal to $-\infty$ and say that the sequence becomes negatively infinite.

We shall have occasion frequently to consider also an infinite sequence of regions of the complex plane so related that each is smaller than the one preceding it and is contained in it. However small one of the regions in the sequence may be, it contains nevertheless an infinite number of points. It is of importance to be able to say when the limit of such a sequence is a single point. This result will be established if it can be shown that every set of points that can be chosen by selecting in any manner whatever a point within or on the boundary of each region has necessarily the same limiting point as the regions are taken smaller and are made to approach zero in area. We shall consider two special cases, namely, a sequence of circles and a sequence of rectangles. In this discussion, we shall make use of a well-known theorem † of the theory of functions of real variables which states that if there is given a definite sequence of intervals defined by

$$I_n = (a_n,\ b_n), \qquad n = 1,\ 2,\ 3,\ \ldots$$

such that the interval I_n lies in I_{n-1} and $\underset{n=\infty}{L} \{Length\ I_n\} = 0$, then if P_n is any point in I_n, end points included, every such set $\{P_n\}$ of points has an unique limit p, and we say that the sequence of intervals $\{I_n\}$ defines the point p.

We shall now consider the following theorem.

THEOREM III. *Let* $\{S_n\} \equiv S_1, S_2, \ldots, S_n, \ldots$ *be an infinite sequence of circles so related that each lies in the preceding one and moreover having the property that*

$$\underset{n=\infty}{L}\ Area\ S_n = 0;$$

then the sequence $\{S_n\}$ *defines a limiting point as n increases indefinitely.*

* See Pierpont, *Theory of Functions of Real Variables*, Vol. I, Arts. 101 and 109.
† *Ibid.*, p. 82.

Let each circle S_n be inclosed in a square by drawing tangent lines parallel to the X-axis and to the Y-axis. We have then a sequence of intervals on each axis (Fig. 9) satisfying the conditions of the foregoing theorem quoted from the theory of functions of real variables. The sequence of intervals

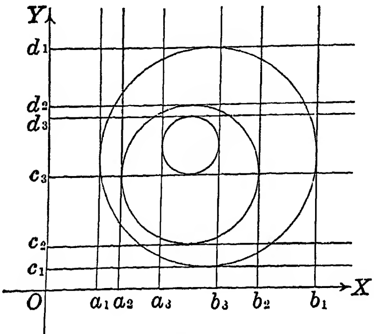

FIG. 9.

$$(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n), \ldots$$

situated on the X-axis defines in the limit a definite point a. Likewise the sequence of intervals

$$(c_1, d_1), (c_2, d_2), \ldots, (c_n, d_n), \ldots$$

on the Y-axis defines in the limit a point b. Drawing through a, b lines parallel to the two axes, respectively, the intersection determines a definite point $a + ib$ of the complex plane, which is the limit of every sequence of points $\{P_n\}$ obtained by taking a point in each region of the sequence $\{S_n\}$. This sequence $\{S_n\}$ therefore defines the point $a + ib$. It will be noted that the demonstration does not preclude the possibility that any or all of the circles S_1, S_2, . . . may be tangent internally.

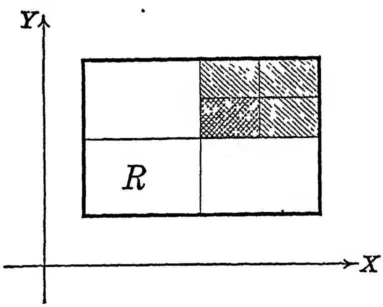

FIG. 10.

In certain discussions, it is more convenient to consider a sequence of rectangles obtained as follows. Suppose we have given the rectangle R, Fig. 10. Divide this rectangle into four equal parts by drawing lines parallel to the two axes. Select one of these four rectangles and divide it into four equal parts in a similar manner. Consider this operation as repeated an indefinite number of times. We may now state the following theorem with reference to the sequence of rectangles obtained.

THEOREM IV. *Given an infinite sequence of rectangles R_1, R_2, . . . , R_n, . . . such that each lies in the preceding one. Let the limit of each dimension of R_n approach zero as n increases without limit. Then the given sequence of rectangles defines in the limit a definite point of the plane.*

The method of proof for this theorem is substantially that given for Theorem III.

By the aid of the foregoing theorem, we readily establish the following theorem.

THEOREM V. *Every infinite set of points in a finite region of the complex plane has at least one limiting point.*

Let R_1 be the rectangle inclosing all of the given points. Divide the rectangle R_1 into four equal rectangles by drawing lines parallel to the sides of R_1. At least one of these smaller rectangles, say R_2, must contain an infinite number of points of the given set. In a similar manner divide R_2 into four equal rectangles; at least one of these rectangles, say R_3, must contain an infinite number of the points of the given set. Regard this method of division as continued indefinitely. We get a sequence of rectangles

$$R_1,\ R_2,\ R_3,\ \ldots,\ R_n,\ \ldots,$$

satisfying the conditions of Theorem IV and hence having a limiting point, say z_0. But as each of the rectangles R_n contains an infinite number of the points of the given set, it follows that z_0 is a limit of the given set. There may or may not be other limiting points, but in every case there must be at least one. The result of the theorem can also be expressed by saying that the given set of points is dense at z_0.

It is frequently desirable to have a condition which is necessary and sufficient for the existence of the limit of a sequence. Such a condition is given by the following theorem.

THEOREM VI. *Given the sequence of complex numbers*

$$\alpha_1,\ \alpha_2,\ \alpha_3,\ \ldots,\ \alpha_n,\ \alpha_{n+1},\ \ldots,\ \alpha_{n+k},\ \alpha_{n+k+1},\ \ldots$$

the necessary and sufficient condition that this sequence has a limit is that corresponding to an arbitrarily small positive number ϵ there exists a positive integer m such that

$$|\alpha_n - \alpha_{n+k}| < \epsilon, \qquad k = 1,\ 2,\ 3,\ \ldots, \qquad n \geqq m.$$

We shall first show that the given condition is necessary. For this purpose suppose α to be the limit of the sequence. If n is taken sufficiently large, say $n \geqq m$, we have from the definition of a limit

$$|\alpha_n - \alpha| < \frac{\epsilon}{2},$$

$$|\alpha - \alpha_{n+k}| < \frac{\epsilon}{2}, \quad k = 1,\ 2,\ 3,\ \ldots,$$

Combining these two inequalities, we have

$$|\alpha_n - \alpha_{n+k}| < \epsilon, \qquad k = 1, 2, 3, \ldots, \qquad n \geqq m$$

which is the condition set forth in the theorem.

The above condition is also sufficient; that is, given the condition

$$|\alpha_n - \alpha_{n+k}| < \epsilon, \qquad k = 1, 2, 3, \ldots, \qquad n \geqq m, \tag{14}$$

it is possible to show that the sequence has a limit. By virtue of this condition we can find among the α's some α_n such that all of the points $\alpha_{n+1}, \alpha_{n+2}, \ldots$ lie within some arbitrarily small distance ϵ from α_n. If we take $\epsilon < \frac{1}{2}$, then it follows from (14) that all of the points α_n, $n \geqq m$, lie within a circle (Fig. 11) having α_m as center and a radius $\frac{1}{2}$. The points α_n, *a fortiori*, lie within the circle C_1 having α_m as center and a radius equal to 1. Among the points α_{m+1}, $\alpha_{m+2}, \ldots$, there exists some one, say α_{m_1}, such that we have

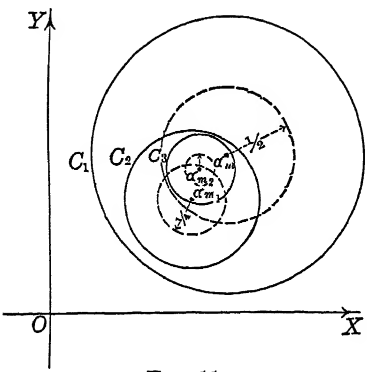

FIG. 11.

$$|\alpha_{m_1} - \alpha_{m_1+k}| < \tfrac{1}{4}, \qquad k = 1, 2, 3, \ldots;$$

that is, all points α_n, $n \geqq m_1$, lie within a distance of $\frac{1}{4}$ from α_{m_1}. These values then, *a fortiori*, lie within the circle C_2 drawn about α_{m_1} as a center with a radius of $\frac{1}{2}$. Among the points α_{m_1+1} $\alpha_{m_1+2}, \ldots$ there can be found one, say m_2, such that

$$|\alpha_{m_2} - \alpha_{m_2+k}| < \tfrac{1}{8}, \qquad k = 1, 2, 3, \ldots.$$

All of these remaining points α_n, $n \geqq m_2$, *a fortiori*, lie within the circle C_3 about the point α_{m_2} having a radius of $\frac{1}{4}$.

It will be observed that C_2 lies within C_1, and C_3 within C_2. Continuing in this manner, we obtain a sequence of circles

$$C_1, C_2, C_3, \ldots, C_n, \ldots$$

fulfilling the conditions of Theorem III; for, each circle lies within the preceding circles and their radii, which are

$$1, \tfrac{1}{2}, \tfrac{1}{4}, \ldots, \tfrac{1}{2^n}, \ldots,$$

respectively, have the limiting value zero. Hence, the sequence of circles defines a definite limiting point, which we may designate by α. The number α is then the limit of the sequence $\alpha_1, \alpha_2, \alpha_3, \ldots$.

13. Continuity. A function of a complex variable is said to be **continuous at a point** if the value of the function at that point is equal to the limit of the values assumed by the function in every neighborhood of the point. There are three things involved in continuity; first, the function must be defined at the point in question; second, the function must have a unique limit as the variable approaches the critical value; and third, the value of the limit must be equal to the value of the function at the point. If either one of the two latter conditions fails, the function is said to be **discontinuous.** If the first condition is not satisfied, then we can not discuss the continuity of the function at the point in question; for, the function does not exist at that point.

This definition does not differ in form from the definition given in calculus for the continuity of a function of a real variable, in which we say that a function $f(x)$ of a real variable x is continuous at $x = a$, if

$$\underset{x \doteq a}{L}\ f(x) = f(a). \tag{1}$$

This definition requires that the same limiting value $f(a)$ be obtained by every possible approach to the point $x = a$, that is, from either the right or the left and through any set of values dense at the point a that may be chosen from those values that x may take in the neighborhood of a. It is necessary that we take into consideration all such values of x in the neighborhood of $x = a$, in determining the existence or non-existence of the limit.

In a similar manner, we say that a function of two real variables $f(x, y)$ is continuous at the point (a, b) with respect to the two variables taken together if we have

$$\underset{\substack{x \doteq a \\ y \doteq b}}{L}\ f(x, y) = f(a, b), \tag{2}$$

which involves the condition that the same limiting value is obtained by all possible methods of approach to the point (a, b) and furthermore, that this limiting value is equal to the value of the function at that point.

The definition of the continuity of a function $f(z)$ of a complex

variable may be briefly expressed by saying that $f(z)$ is continuous at an inner point $z = \alpha$ of a region S, if we have

$$\underset{z=\alpha}{L} f(z) = f(\alpha). \tag{3}$$

More is involved in this definition than in the corresponding definition for continuity of a function of a single real variable. The variable x has but one degree of freedom; that is, it can vary along the real axis only. It can approach the limiting position from two possible directions. On the other hand, the variable $z = x + iy$ can be said to have two degrees of freedom since x and y are independent variables. The variable z can then approach its limiting position, not only from two possible directions, but from any direction in the plane, or through any set of values of z dense at α. In order to affirm that $f(z)$ is continuous, we must be able to say that the same limiting value, namely $f(\alpha)$, is obtained if z is allowed to approach α through every possible set of points dense at α.

If α is a boundary point of a region, then $f(z)$ is said to be continuous at α if $f(z)$ approaches $f(\alpha)$ through every set of inner points of the region dense at α. We say that a function is continuous throughout a region, whether open or closed, if it is continuous at each point of the region.

We can establish the following theorem as a consequence of the definition of continuity.

THEOREM I. *If $f(z)$ is continuous at an inner point z_0 of a region S and if $f(z_0) \neq 0$, then there exists a neighborhood of z_0 for which $f(z) \neq 0$.*

Since $f(z)$ is continuous at $z = z_0$, we have

$$\underset{z \doteq z_0}{L} f(z) = f(z_0);$$

that is, for an arbitrarily small positive number ϵ, there exists a positive number δ such that

$$|f(z) - f(z_0)| < \epsilon \tag{4}$$

for $|z - z_0| < \delta$. But as $f(z_0)$ is a constant different from zero, we may write

$$|f(z_0) - 0| = A > 0. \tag{5}$$

By taking $\epsilon < \frac{A}{2}$, we have by combining (4) and (5)

$$|f(z) - 0| > \frac{A}{2}, \qquad |z - z_0| < \delta.$$

But as A is greater than zero, this relation establishes the theorem.

The continuity of $f(z) = u + iv$ with respect to z may be seen to depend upon that of $u(x, y)$, $v(x, y)$ with respect to x, y, as stated in the following theorem.

THEOREM II. *The necessary and sufficient condition that $f(z)$ is continuous at $z = \alpha$ is that $u(x, y)$, $v(x, y)$ are both continuous in x, y at the corresponding point (a, b), where $\alpha = a + ib$.*

This result follows as a direct consequence of Theorem II of Art. 12 and the definition of continuity.

It has already been pointed out that when $f(z)$ is continuous at $z = \alpha$, we may write

$$|f(z) - f(\alpha)| < \epsilon, \qquad |z - \alpha| < \delta, \tag{6}$$

where ϵ is an arbitrarily small positive number and δ is another positive number depending for its value upon ϵ and α. In other words, ϵ is first selected as small as we choose and then δ is so determined that the condition given in (6) is satisfied. For any fixed value of ϵ, the value of δ may change when some point other than $z = \alpha$ is taken as the limiting point. Moreover, for any particular value of z, various values may be assigned to δ. If z_0 is any point in a given region, denote by $\delta(z_0)$ the value of δ corresponding to the previously assigned value of ϵ. Then, for this given ϵ, $\delta(z)$ is a function of z, made up of the totality of all the values of $\delta(z)$ as z takes the various values in the given region.

Consider now the function $\delta(z)$ for all values of z in the given region. If for an arbitrarily chosen $\epsilon > 0$, $\delta(z)$ has a lower limit δ' different from zero, we say that the function $f(z)$ is **uniformly continuous** in the given region. From what has been said, it will now be seen that an essential characteristic of uniform continuity is that the relation (6) is satisfied by any definite value δ, where $0 < \delta < \delta'$, regardless of the value of α.

It is shown in the theory of functions of a real variable that any function that is continuous throughout a closed interval is uniformly continuous in that interval. The corresponding theorem for complex variables may be stated as follows:

THEOREM III. *If a function $f(z)$ of the complex variable z is continuous in a finite closed region S, then it is uniformly continuous in that region.*

To prove this proposition, we shall assume the contrary to be true and show that this assumption leads to a contradiction. The

assumption is then that the function $f(z)$ does not satisfy the definition for uniform continuity in the closed region S. Since $f(z)$ is continuous at every point within S or upon its boundary, we know from the foregoing discussion that for a fixed but previously assigned value of ϵ there is associated with each point z of the region a definite number $\delta(z)$, which, however, may vary with the point. The function $\delta(z)$ is fully defined at all points within S or upon its boundary. By the assumption that $f(z)$ is not uniformly continuous, we have the condition that the lower limit of $\delta(z)$ is zero. Inclose the region S in a rectangle by drawing lines parallel to the two axes.

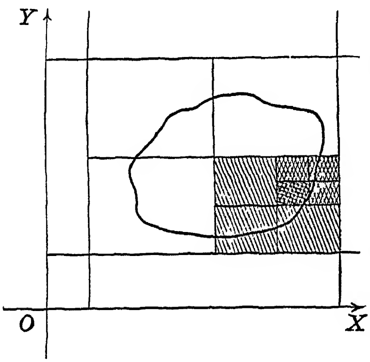

FIG. 12.

Divide this rectangle R_1 into four equal parts by again drawing lines parallel to the axes. In the part of S lying in at least one of these subdivisions, say R_2, $\delta(z)$ must have the lower limit zero. Divide in the same way R_2 into four equal parts; in one of these divisions, say R_3, $\delta(z)$ has the lower limit zero. Continue this process indefinitely. In the limit the sequence of rectangles $R_1, R_2, R_3, \ldots$ defines a definite point z' (Art. 12), which may be a point within or upon the boundary of the region S. In any case, since S is a closed region, z' is a point of S. We may then say that there is, under the assumption as to uniform continuity, at least one point z' of the given region such that in every neighborhood of z' the lower limit of the δ's is zero.

The given function $f(z)$ is, however, continuous for $z = z'$, and hence for the point z' there exists a δ_0 different from zero, where $\delta_0 \leqq \delta(z')$, such that for any two values z_1, z_2 of the variable, for which

$$|z_1 - z'| < \delta_0, \qquad |z_2 - z'| < \delta_0,$$

we have

$$|f(z_1) - f(z')| < \frac{\epsilon}{2},$$

$$|f(z_2) - f(z')| < \frac{\epsilon}{2}.$$

Combining these inequalities, we have

$$|f(z_1) - f(z_2)| < \epsilon;$$

that is, the oscillation of the function within a circle of radius δ_0 about z' can not exceed the arbitrarily small number ϵ. If we now draw another circle C of radius $\frac{\delta_0}{2}$ about z' as a center, then at every point within C the given function $f(z)$ is not only continuous but if a circle of radius $\frac{\delta_0}{2}$ be drawn about any such point the upper limit of the oscillation of $f(z)$ within this circle likewise can not exceed the arbitrarily small number ϵ. Hence, the value of δ associated with any point in C can not be less than $\frac{\delta_0}{2}$, which in turn is greater than zero. If the point z' lies upon the boundary of S, we consider only that portion of the region bounded by the two circles which lies in S. Consequently, we can now say that the lower limit of the δ's for all points lying within C can not be zero as assumed. From this contradiction the theorem follows.

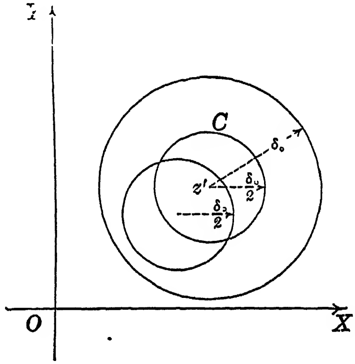

FIG. 13.

From the definition of continuity of $f(z)$ at a boundary point and from the foregoing theorem, we conclude that if $f(z)$ is continuous at each point of an arc C, end points included, of the boundary of a closed region S, then we have for any point z_0 of C

$$| f(z) - f(z_0) | < \epsilon, \qquad | z - z_0 | < \delta,$$

where the values of z correspond to inner points of S and δ is independent of z_0; that is, for a given ϵ there exists a δ that satisfies the required conditions equally well for all points z_0 of C, end points included. We say then that $f(z)$ **converges uniformly** along C. Expressed in terms of limits, it follows that $f(z)$ converges uniformly along the arc C if the limit

$$\underset{z \doteq z_0}{L} f(z) = f(z_0)$$

exists when taken over inner points of S for each point z_0 of C, end points included.

We are able now to state the following theorem concerning uniform convergence along an arc.

THEOREM IV. *If $f(z)$ is defined for a closed region S and converges uniformly along an arc C of the boundary of S, then $f(t)$ is continuous, where t denotes the values of z on C.*

Let t_0 be any point of the arc C. Then from the definition of uniform convergence, we have

$$| f(z) - f(t_0) | < \epsilon, \qquad | z - t_0 | < \delta, \tag{1}$$

where z is confined to inner points of S, that is to inner points within a circle γ_0 of radius δ about t_0. Let t_1 be any other point of C situated within γ_0. Then by hypothesis, we have also

$$| f(z) - f(t_1) | < \epsilon, \qquad | z - t_1 | < \delta \tag{2}$$

for all inner points of S within a circle γ_1 of radius δ about t_1. The two circles overlap and inclose common inner points of S. For such a common point z' of z we have from (1)

$$| f(z') - f(t_0) | < \epsilon,$$

and likewise from (2), we get

$$| f(z') - f(t_1) | < \epsilon.$$

Combining these two inequalities, we obtain

$$| f(t_1) - f(t_0) | < 2\,\epsilon.$$

But t_1 is any point of C such that $| t_1 - t_0 | < \delta$. Hence, $f(t)$ is continuous at t_0 and therefore for all values of t along C, as stated in the theorem.

THEOREM V. *If $f(z)$ is continuous in a finite closed region S, then there exists some finite number M such that $|f(z)| < M$ for all values of z in S.*

To establish this theorem, we assume the contrary to be true, namely, that there exists no finite number M answering the conditions of the theorem, and shall prove that this assumption leads to a contradiction of the given hypothesis. Inclose the given region in a rectangle by drawing lines parallel to the two axes and divide this rectangle into four equal parts. In at least one of these subdivisions $|f(z)|$ exceeds every finite value. Let R_1 be such a subdivision. Divide R_1 into four equal parts likewise by drawing lines parallel to the two axes. In at least one of these new subdivisions, say R_2, $|f(z)|$ must also exceed every finite bound. Divide R_2 into four

equal parts in the same manner and regard this process as continued indefinitely. In this way a sequence of rectangles

$$R_1, R_2, \ldots, R_n, \ldots$$

is obtained satisfying the conditions of Theorem IV, Art. 12.

In the limit these rectangles define a point, say z_0, and as the given region S is a closed region the point z_0 is a point of S. It can now be said that in every deleted neighborhood of z_0, $f(z)$ exceeds in absolute value every finite bound. Consequently, the limit $\underset{z \doteq z_0}{L} f(z)$ can not be said to exist; because, a set of points $z_1, z_2, z_3, \ldots, z_n, \ldots$, having z_0 as a limiting point, can be selected so that as z approaches z_0 through these points, the function $|f(z)|$ exceeds all finite limits, that is, becomes infinite. From the definition of continuity, it follows then that $f(z)$ is discontinuous at z_0. This conclusion is a contradiction of the hypothesis set forth in the theorem and from this contradiction the theorem follows.

THEOREM VI. *If $f(z)$ is continuous in a finite closed region S, then $|f(z)|$ has a finite upper limit in S.*

From Theorem V we know that $|f(z)|$ has a finite upper bound; that is, there exists a finite number M such that $|f(z)| < M$. The function $f(z)$ is then represented by points lying within a circle C_1 about the origin having the radius M. The present theorem asserts that there exists a circle of radius equal to or less than M such that there are values of $f(z)$ represented by points indefinitely close to or upon this circle but not without it. Divide the radius M of the circle C_1 into 10 equal parts by drawing about the origin concentric circles having the radii

$$\frac{M}{10}, \frac{2M}{10}, \ldots, \frac{9M}{10}.$$

Among these circles, including C_1, there is a smallest one such that no values of $f(z)$ are represented by points lying outside of it. Suppose the next smaller circle C_2 has the radius $\frac{k_1M}{10}$, where k_1 may have any one of the values $0, 1, 2, \ldots, 9$. If $k_1 = 0$, the circle C_2 becomes a point, namely the origin. Between C_2 and the next larger circle insert 10 other concentric circles having respectively the radii

$$\frac{k_1M}{10} + \frac{M}{10^2}, \frac{k_1M}{10} + \frac{2M}{10^2}, \ldots, \frac{k_1M}{10} + \frac{9M}{10^2}.$$

Among these circles, including the circle of radius $\dfrac{(k_1+1)\,M}{10}$, there is likewise a smallest such that no values of $f(z)$ are represented by points exterior to it. Let the circle next smaller than that circle be C_3, having the radius $\dfrac{k_1M}{10}+\dfrac{k_2M}{10^2}$, where again k_2 may take any one of the values 0, 1, 2, . . . , 9. Continuing in this manner, we get a sequence of circles

$$C_2,\ C_3,\ C_4,\ \ldots,$$

having respectively the radii

$$\frac{k_1M}{10},\quad \frac{k_1M}{10}+\frac{k_2M}{10^2},\quad \frac{k_1M}{10}+\frac{k_2M}{10^2}+\frac{k_3M}{10^3},\ \ldots.$$

This sequence of numbers satisfies the conditions of Theorem VI, Art. 12, and hence has a limit, say G. This limit is, however, the upper limit required, for the sequence determines the radius of the least circle such that no values of $f(z)$ are represented by points without it.

We may now state the following theorem:

THEOREM VII. *If $f(z)$ is continuous in a finite closed region S, then $|f(z)|$ attains its upper limit in S.*

This theorem is equivalent to saying that every function $f(z)$ which is continuous throughout a finite closed region is such that $|f(z)|$ possesses a maximum value.

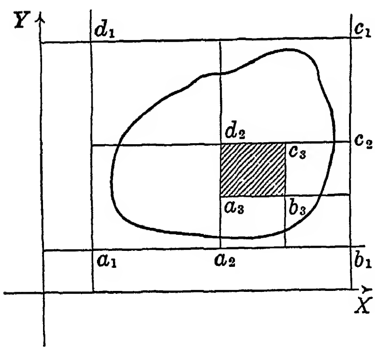

FIG. 14.

Construct the rectangle (a_1, b_1, c_1, d_1), Fig. 14, inclosing the given region S for which the function is defined. Divide this rectangle into four equal parts by drawing lines parallel to the two axes. Consider only those rectangles that contain at least one point of the given region S. By Theorem VI, $|f(z)|$ has in S a finite upper limit. Denote this upper limit by G. Then in some one of the subdivisions of the original rectangle, say (a_2, b_1, c_2, d_2), $|f(z)|$ must have the upper limit G. Divide the rectangle (a_2, b_1, c_2, d_2) likewise into four equal parts as shown in the figure; some one of these new divisions, say (a_3, b_3, c_3, d_2) must contain such values of $f(z)$ that $|f(z)|$ has the

upper limit G. Consider this process of subdivision as continued indefinitely. By each division, some one of the subdivisions must be such that $|f(z)|$ has the upper limit G for the values of z in it. By this method of subdivision each rectangle that is chosen is situated within all of those that have entered previously into consideration. The dimensions of the rectangles have the limit zero and the sequence of rectangles defines in the limit a definite point, which we may denote by z_0. Since $f(z)$ is continuous for $z = z_0$, we have

$$\underset{z \doteq z_0}{L} f(z) = f(z_0). \tag{1}$$

But as the limit of the absolute values of a function is the absolute value of the limit, we have also

$$\underset{z \doteq z_0}{L} |f(z)| = |f(z_0)|. \tag{2}$$

In every neighborhood of z_0 the upper limit of $|f(z)|$ is G. The limit of $|f(z)|$, as z approaches z_0, can not then be greater than G. We shall now show that this limit can not be less than G; for, suppose it is $G - A$, where $A \neq 0$. We then have for an arbitrarily chosen positive number ϵ, say $\epsilon \leqq \dfrac{A}{2}$, another positive number δ such that

$$\big|\, |f(z)| - (G - A) \,\big| < \epsilon, \qquad |z - z_0| < \delta;$$

that is, for all values of z within the circle having z_0 as center and δ as radius $|f(z)|$ differs from $G - A$ by less than ϵ. This is a contradiction to the foregoing conclusion that the upper limit of $f(z)$ in every neighborhood of z_0 is G. Hence, the limit $\underset{z \doteq z_0}{L} |f(z)|$ can not be less than G, and as it must exist and can not exceed G, it must be equal to G. We have then

$$\underset{z \doteq z_0}{L} |f(z)| = G. \tag{3}$$

By comparing (2) and (3), we obtain

$$|f(z_0)| = G,$$

and the theorem is established.

EXERCISES

1. Classify the following functions:

$$
\begin{aligned}
&(a)\ 3x^3 + 7x^2 + 19,\\
&(b)\ 5x^{\frac{2}{3}} + 2x^{\frac{1}{2}} + 6x + 3,\\
&(c)\ 2x^{-3} + 7x^2 + 3x^{-1} + 4,\\
&(d)\ \tan x,\ \cos x,\ \log x.
\end{aligned}
$$

2. Show that 1 is the limit of the sequence

$$\tfrac{2}{3},\ \tfrac{4}{5},\ \tfrac{6}{7},\ \tfrac{8}{9},\ \ldots.$$

3. Given $f(z) = (x^4 + y^4 - 6x^2y^2) + i(4x^3y - 4xy^3)$. Find the limit of $f(z)$ as z approaches $2 + 3i$.

4. Show that the sequence of circles given by the equation

$$\left(x - \frac{1}{n}\right)^2 + y^2 = \frac{1}{n^2}, \qquad n = 1, 2, 3, \ldots$$

defines in the limit one and only one point, namely the origin.

5. Show that $f(z) = \log(x^2 + y^2) + i \operatorname{arc\,tan} \dfrac{2xy}{x^2 - y^2}$ is continuous for finite values of $z \equiv x + iy \neq 0$.

6. Show that an integral rational function of z is uniformly continuous in any given finite region.

7. Given $f(z) = \dfrac{z^2 + 1}{z^2 - 1}$. Show that $|f(z)|$ has a finite upper limit in the region bounded by the circle $x^2 + y^2 = \frac{1}{4}$. Find this upper limit.

8. Find the limit of the sequence

$$\alpha_1\ \alpha_2,\ \ldots\ \alpha_n,\ \ldots$$

where

$$\alpha_n \equiv \frac{1}{2^n} + i\left[1 - \left(\frac{3}{4}\right)^n\right].$$

9. Given a closed region S in which $f(z)$ is continuous. Show that for an arbitrarily chosen positive number ϵ there exists another positive number δ such that

$$|f(z_1) - f(z_2)| < \epsilon,$$

for $|z_1 - z_2| < \delta$, z_1, z_2 being points of S.

CHAPTER III

DIFFERENTIATION AND INTEGRATION

14. Differentiation; analytic function. The definition of the derivative of a function $f(z)$ of a complex variable is identical in form with the definition of the derivative of a function of a real variable. Let z be a variable point in the neighborhood of z_0. Put

$$\Delta z \equiv z - z_0,$$

and

$$\Delta w \equiv f(z) - f(z_0) = f(z_0 + \Delta z) - f(z_0).$$

If the limit

$$\underset{\Delta z \doteq 0}{L} \frac{\Delta w}{\Delta z} \equiv \underset{\Delta z = 0}{L} \frac{f(z_0 + \Delta z) - f(z_0)}{\Delta z}$$

exists, we say that this limit is the **derivative** of $f(z)$ at the point z_0. The derivative is then the limit of the ratio of the two complex variables Δw and Δz. From the properties of limits already discussed, it will be seen that to have this limit exist, the same limiting value must be obtained independently of the path along which z approaches z_0. Moreover, the same limiting value must be obtained if we select any possible set of values for z dense at z_0 and let z approach z_0 through these values. The value of the limit must therefore be independent of the amplitude of Δz as z approaches z_0. We shall make use of the same symbols as in the calculus of a real variable to denote a derivative; for example, the derivative of $w = f(z)$ with respect to z is denoted by any one of the symbols:

$$D_z w, \quad \frac{dw}{dz}, \quad f'(z).$$

The general laws of differentiation for real variables can be extended without modification to functions of a complex variable, since they depend upon the general laws of limits, which hold equally well in both fields. For example, we have

$$(a) \qquad D_z(cw) = cD_z w,$$
$$(b) \qquad D_z(w_1 + w_2) = D_z w_1 + D_z w_2,$$
$$(c) \qquad D_z(w_1 \cdot w_2) = w_1 D_z w_2 + w_2 D_z w_1,$$

(d) $$D_z\left(\frac{w_1}{w_2}\right) = \frac{w_2 D_z w_1 - w_1 D_z w_2}{w_2^2},$$

(e) $$D_z f(w) = D_w f(w) \cdot D_z(w), \text{ etc.}$$

It is suggested that the student deduce the foregoing laws for differentiation directly from the definition of a derivative. As with functions of a real variable, the continuity of the function is a necessary condition for the existence of the derivative.

The differentials dw, dz may be defined in precisely the same manner that differentials are defined in the calculus of real variables. For this purpose, suppose w and z are expressed in terms of a third common variable t. We have then

$$w = w(t), \qquad z = z(t), \qquad \text{where} \quad w = f(z).$$

If we differentiate with respect to the common variable t, we have

$$D_t w = D_z f(z) \cdot D_t z.$$

As the derivatives $D_t w$, $D_t z$ with respect to the common variable enter homogeneously into the above identity, we define the differentials dw, dz as numbers equal to or in proportion to these derivatives and write

$$dw = D_z f(z)\, dz.$$

The parametric representation introduced above gives us some advantage in certain discussions. For example, we shall frequently have occasion to introduce the condition that

$$x = \phi(t), \qquad y = \psi(t),$$

where ϕ, ψ are real functions of a real variable t and possess continuous first derivatives with respect to t. We then have from the definition of a differential

$$dx = \phi'(t)\, dt, \qquad dy = \psi'(t)\, dt. \tag{1}$$

The connection between the real functions $\phi(t)$, $\psi(t)$ and the complex variable z is given by the equation $z = x + iy$. We have therefore

$$\begin{aligned} dz &= D_t z\, dt \\ &= \{\phi'(t) + i\,\psi'(t)\}\, dt. \end{aligned}$$

Replacing $\phi'(t)\, dt$, $\psi'(t)\, dt$ by their values as given in (1) we have

$$dz = dx + i\, dy.$$

The higher derivatives and higher differentials follow the same laws as in the calculus of real variables and the same symbols are used to represent them.

As we have already pointed out (Art. 11), the general definition of a function does not impose upon the functional correspondence such special properties as continuity, differentiability, etc. To say that $f(z)$ is a function of the complex variable z asserts nothing further than that $f(z)$ depends upon z in such a manner that for each value given to z there is thereby determined a definite value or set of values of the function $f(z)$. We make a substantial advance when we can ascribe to a function the properties of continuity and differentiability. The functions to be considered in this volume possess for the most part both of these properties.

If a given single-valued function $f(z)$ has a uniquely determined derivative at the point α and at every point in the neighborhood of α, then $z = \alpha$ is called a **regular point** of $f(z)$. By some authors, the function is said to be analytic at $z = \alpha$ and by others it is called holomorphic at this point. We shall, however, reserve these terms for other uses.

A point in every deleted neighborhood of which there are regular points but which is itself not a regular point is called a **singular point** of the given function.

If every point of a given region S is a regular point of a single-valued function $f(z)$, then $f(z)$ is said to be **holomorphic** in S. It should be borne in mind throughout this and the succeeding chapters that we have defined a region to be a continuum of inner points; hence, it is understood that a region does not include its boundary points unless so specified. We shall speak of a function $f(z)$ as being an **analytic function** of z if it is holomorphic in at least some region S with the possible exception of certain singular points which do not interrupt the continuity of S. It is always possible then to join any two regular points of S by a continuous curve which lies wholly within S and which does not pass through a singular point. A more precise definition of analytic functions will be given in Chapter VII.

From the definition of a function which is holomorphic in a region, we have at once the following general properties. Given two functions $f(z)$, $\phi(z)$, each holomorphic in a region S; then it follows that in S:

1. $f(z) + \phi(z)$ *is holomorphic,*
2. $f(z) \cdot \phi(z)$ *is holomorphic,*

3. $\dfrac{f(z)}{\phi(z)}$ *is holomorphic, except for those values of z for which $\phi(z)=0$.*

4. *If w_0 is a regular point of $f(w)$, and z_0 is a regular point of $w = \phi(z)$, where $\phi(z_0) = w_0$, then z_0 is a regular point of the function $f\{\phi(z)\}$ considered as a function of z.*

From these properties, it follows that every rational integral function of z is an analytic function, holomorphic in the finite region of the complex plane. Since every rational function is holomorphic, except at most at a finite number of points where the denominator is zero, it also is an analytic function.

15. Line-integrals. It was pointed out in the last article that the definition of the derivative of a function of a complex variable involves a more complicated limit than the corresponding definition in the case of functions of a real variable. A similar generalization is necessary in the discussion of integration, in that we must in general take into account the path along which the integral is to be taken. In the case of functions of a real variable, the independent variable x can pass continuously from any value x_1 to some other value x_2 along only one path, namely, by passing through the intermediate values along the X-axis. In the case of functions of a complex variable, the independent variable z can pass from a value z_1 to another value z_2 by any number of different paths. Consequently, the definition of a definite integral between two values of z can have a significance only when we consider the path by which z passes from the one value to the other.

As the subject is not always considered in elementary text-books on calculus, we shall now define a **line-integral** and discuss some of the more general properties of such integrals. Among other things, we shall show that the integral of a function of a complex variable taken over a given path may be expressed in terms of line-integrals of functions of the real variables x, y taken over the same path.

In the calculus of real variables a definite integral is defined as the limit of a sum; that is

$$\int_a^b f(x)\,dx = \underset{n=\infty}{L} \sum_{k=1}^{n} f(\xi_k)\,\Delta_k x. \tag{1}$$

Stated in words: the portion of the X-axis between a and b is divided into n parts, the length $\Delta_k x$ of each division is multiplied by the value of the function at some arbitrary point ξ_k in that division and the limit of the sum of these products is taken, as the number of such divi-

sions increases without limit while the length of the divisions simultaneously approaches zero. It is of importance to observe that the n divisions between the limits of integration are taken along the X-axis and each of these divisions is multiplied by a value of the function at a point on the same axis. Suppose instead, these divisions and the points at which the functional values are to be used as multipliers are taken along some curve C, called the **path of integration,** the function with whose values we are concerned now being a function of the two variables x, y. Let the functional value at a point on the curve in each division be multiplied by Δx, which is the orthogonal projection upon the X-axis of the division of the curve. The limit of the sum of these products as Δx approaches zero, that is as the number of divisions increases indefinitely, is a line-integral of the given function along the path C. This curve may lie in the XY-plane, or in case we have a function of three real variables, the path of integration may be a curve in space. In the particular applications to be made of integrals in the present volume the path of integration will always be a plane curve. Any rectifiable curve, that is any curve having a definite length, may be taken as the path of integration. However, as there is a certain element of arbitrariness in the choice of the path of integration, we shall avoid certain complications in the discussions to follow by taking as that path a curve that may be broken up into a finite number of divisions, each of which is either a rectilinear segment parallel to one of the coördinate axes or else has the property that it is determined by a function $y = \phi(x)$, where $\phi(x)$ and its inverse function $x = \psi(y)$ are single-valued and have first derivatives that are continuous except at most at the end points, at which points they may become infinite. Such a curve is monotone by segments and for convenience will be designated as an **ordinary curve,*** whenever a special name is necessary for the sake of clearness. However, in the present volume only ordinary curves will be employed.

Let AB be one of the finite number of divisions of which an ordinary curve C (Fig. 15) is composed. Let the coördinates of the points A, B be (x_0, y_0) and (x_n, y_n), respectively. Divide the arc AB into n parts by the insertion of $n - 1$ points $p_1, p_2 \ldots, p_k \ldots, p_{n-1}$, whose coördinates are $(x_1, y_1), (x_2, y_2), \ldots (x_k, y_k), \ldots (x_{n-1}, y_{n-1})$, respectively.

* Compare: Pringsheim, *Encyklopädie der Math. Wiss.*, II, A 1, p. 22; also Dodd, *Bull. of Univ. of Tex.*, No. 222, March, 1912.

Select at pleasure a point (ξ_k, η_k) upon each arc (p_{k-1}, p_k), $k = 1, 2, \ldots, n$. Having given a function $F(x, y)$ which is continuous in x and y together along the curve C, form the sum of the products of the subintervals

$$\Delta_k x \equiv x_k - x_{k-1}$$

and the values of the given function $F(x, y)$ at the points (ξ_k, η_k). Finally, consider the limit of this sum as the number of subintervals $\Delta_k x$ between A and B is increased indefinitely, while at the same time the length of each interval approaches zero, namely, the limit

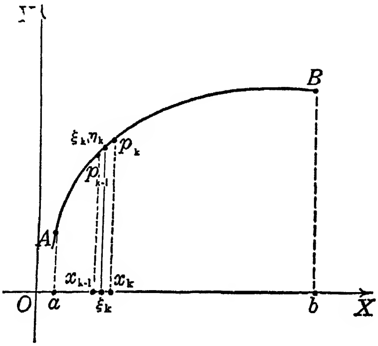

FIG. 15.

$$\underset{n=\infty}{L} \sum_{k=1}^{n} F(\xi_k, \eta_k) \Delta_k x. \qquad (2)$$

If this limit exists, it is called a line-integral * of $F(x, y)$ along the given curve C between the limits A and B. Such an integral is represented by the symbols

$$\int_C F(x, y)\, dx, \qquad \int_{x_0, y_0}^{x_n, y_n} F(x, y)\, dx, \qquad \int_{AB} F(x, y)\, dx.$$

In defining a line-integral, we took the limit of a sum of products formed by multiplying $F(\xi_k, \eta_k)$ by the projection of the arc (p_{k-1}, p_k) upon the X-axis. By taking the projection of this arc upon the Y-axis, we may define in an analogous manner the line-integral

$$\int_C F(x, y)\, dy.$$

It will be observed that the ordinary definite integral is merely a special case of a line-integral, namely, where one of the axes of coordinates is taken as the path of integration.

The existence of the limit defining a line-integral may be made to depend upon that defining an ordinary definite integral. Let us assume that $F(x, y)$ is a continuous function of the two variables x, y together along the path of integration. Let an arc AB of the curve $y = \phi(x)$ be selected as the path of integration (Fig. 16). For the present, we shall also restrict the discussion to the case where

* Sometimes called also a curvilinear integral.

no line parallel to the Y-axis cuts this arc in more than one point. We may replace y by $\phi(x)$ and write

$$F(x, y) = F(x, \phi(x)) = f(x), \tag{3}$$

where $f(x)$ is a continuous function. The limit considered in (2) then becomes

$$\underset{n=\infty}{L} \sum_{k=1}^{n} f(\xi_k)\, \Delta_k x.$$

Since $f(x)$ is a continuous function, this limit exists and defines the definite integral* $\int_a^b f(x)\, dx$, and we have

$$\int_a^b f(x)\, dx = \int_{AB} F(x, \phi(x))\, dx, \tag{4}$$

where a, b are the projections of A, B, respectively, upon the X-axis.

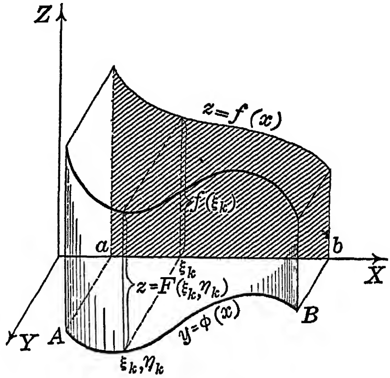

FIG. 16.

Consequently, the line-integral $\int_C F(x, y)\, dx$ not only exists when $F(x, y)$ is continuous in x, y, but we may write

$$\int_C F(x, y)\, dx = \int_a^b f(x)\, dx. \tag{5}$$

It will be observed that the integral $\int_C F(x, y)\, dx$ depends in general upon the curve C as well as upon the function $F(x, y)$; for, taking x, y, z as the space coördinates of a point, the integral $\int_a^b f(x)\, dx$ is represented by the shaded area (Fig. 16) under the curve $z = f(x)$. This area is the projection upon the XZ-plane of the area upon the cylinder perpendicular to the XY-plane through the path of integration $y = \phi(x)$, and underneath the curve of intersection of this cylinder and the surface $z = F(x, y)$. As the path of integration $y = \phi(x)$ changes, the cylinder changes and of course the projected area may change.

In the discussion thus far we have considered only the case where the curve $y = \phi(x)$ is cut by a line parallel to the Y-axis in but a

* See Townsend and Goodenough, *First Course in Calculus*, p. 177, Art. 80.

single point. It may happen that such a line may meet the given arc AB in more than one point, say at the points y_1, y_2, . . . , as shown in Fig. 17. In such a case the given arc should be divided into several portions such that each portion satisfies the required condition. As the path of integration is an arc of an ordinary curve, the number of such subdivisions is always finite. In the case shown in the figure, the arc AB may be decomposed into the arcs AD, DE, EB where each satisfies the necessary condition. We may therefore write

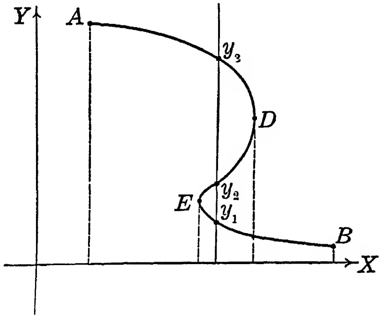

FIG. 17.

$$\int_{AB} F(x, y)\, dx = \int_{AD} F(x, y)\, dx + \int_{DE} F(x, y)\, dx + \int_{EB} F(x, y)\, dx. \quad (6)$$

If P, Q are two real functions of x, y, we shall understand by the line-integral

$$\int_{x_0,\, y_0}^{x_n,\, y_n} P\, dx + Q\, dy, \quad \text{or} \quad \int_C P\, dx + Q\, dy, \qquad (7)$$

the sum of the two line-integrals

$$\int_{x_0,\, y_0}^{x_n,\, y_n} P\, dx, \qquad \int_{x_0,\, y_0}^{x_n,\, y_n} Q\, dy.$$

From what has been said, it follows that these integrals exist if along C the functions P, Q are continuous in x, y taken together.

It is frequently convenient to change the independent variables x, y so as to express the equation of the path of integration in a parametric form. For example, suppose we have

$$x = \Psi_1(t), \qquad y = \Psi_2(t), \qquad (8)$$

where $\Psi_1(t)$, $\Psi_2(t)$ are continuous functions of the real variable t having continuous single-valued first derivatives.

As the point (x, y) varies from A to B along the given path of integration, suppose t varies from t_0 to t_n. Corresponding to the divisions (p_{k-1}, p_k) of the arc AB, we have the increments $\Delta_k t = t_k - t_{k-1}$, $k = 1, 2, \ldots, n$. By the law of the mean, we have then from (8)

$$x_k - x_{k-1} = \Psi_1'(t_k') \cdot (t_k - t_{k-1}),$$

where t_k' lies between t_{k-1} and t_k. Corresponding to t_k' there is a point (ξ_k, η_k) of the arc (p_{k-1}, p_k). Hence, if we have given a continuous function $P(x, y)$, we may write

$$\sum_{k=1}^{n} P(\xi_k, \eta_k)\, \Delta_k x = \sum_{k=1}^{n} P\{\Psi_1(t_k'),\ \Psi_2(t_k')\}\, \Psi_1'(t_k')\, \Delta_k t.$$

Passing to the limit as n becomes infinite, we have from the definition of a line-integral

$$\int_{AB} P(x, y)\, dx = \int_{t_0}^{t_n} P\{\Psi_1(t),\ \Psi_2(t)\}\Psi_1'(t)\, dt. \tag{9}$$

In a similar manner, we may show that if $Q(x, y)$ is continuous in x, y together, we have

$$\int_{AB} Q(x, y)\, dy = \int_{t_0}^{t_n} Q\{\Psi_1(t),\ \Psi_2(t)\}\Psi_2'(t)\, dt. \tag{10}$$

For the general form of the line-integral as given in (7), we have then

$$\int_{AB} P\, dx + Q\, dy = \int_{t_0}^{t_n} \{P \cdot \Psi_1'(t) + Q \cdot \Psi_2'(t)\}\, dt. \tag{11}$$

The integrals in the second members of (9), (10), (11) are ordinary definite integrals. From the relations expressed in these equations, the laws of operation with line-integrals may be deduced from those of ordinary definite integrals. The following consequences of these relations are to be especially noted.

1. *The law for the change of variable in ordinary definite integrals applies likewise to the more general case of line-integrals.*

2. *The integrals*

$$\int_{AB} P\, dx + Q\, dy, \qquad \int_{BA} P\, dx + Q\, dy$$

have the same numerical value, but are opposite in sign.

3. *If z_0 is any point upon the path of integration AB, then*

$$\int_{AB} P\, dx + Q\, dy = \int_{Az_0} P\, dx + Q\, dy + \int_{z_0B} P\, dx + Q\, dy.$$

The function to be integrated may involve a parameter in addition to the variables x, y. It is sometimes desirable to be able to differ-

entiate such an integral with respect to the parameter. Suppose we have given the line-integral

$$\int_{AB} P(x, y, a)\, dx,$$

where $P(x, y, a)$ is continuous in x, y, a taken together and where the path of integration AB is independent of a. Suppose also that $\frac{\partial P}{\partial a}$ exists and is likewise continuous in x, y, a. We may then show that

$$\textbf{4.} \qquad \frac{d}{da}\int_{AB} P(x, y, a)\, dx = \int_{AB} \frac{\partial P(x, y, a)}{\partial a}\, dx. \tag{12}$$

This result follows as a consequence of (9). We have

$$\frac{d}{da}\int_{AB} P(x, y, a)\, dx = \frac{d}{da}\int_{t_0}^{t_n} P\{\Psi_1(t),\ \Psi_2(t),\ a\}\, \Psi_1'(t)\, dt. \tag{13}$$

Since $\frac{\partial P}{\partial a}$ exists and is continuous in (t, a), we have *

$$\frac{d}{da}\int_{t_0}^{t_n} P\{\Psi_1(t),\ \Psi_2(t),\ a\}\, \Psi_1'(t)\, dt = \int_{t_0}^{t_n} \frac{\partial P\{\Psi_1(t), \Psi_2(t), a\}}{\partial a}\, \Psi_1'(t)\, dt. \tag{14}$$

This last integral is by (9) equal to $\int_{AB} \frac{\partial P(x, y, a)}{\partial a}\, dx.$

Hence, we have from (13) and (14) the required relation as stated in (12).

The path of integration may be a closed curve, that is, it may be the boundary of a given region, in which case the limits of integration are represented by the same point of the plane. There is still a choice of direction in which the integral is to be taken. We say that it is taken in a **positive sense** with respect to the region bounded, if it is so taken that this region lies always to the left of the observer as he proceeds along the curve in the direction in which the integral is taken.

Often the boundary of a region consists of two or more closed curves. For example, the region S in Fig. 18 is bounded by the curve M and the circles 1, 2, 3. We may speak of M as the outer portion of the boundary and of the circles 1, 2, 3 as inner portions of the boundary. A region is said to be **simply connected** if every closed curve in it forms by itself a complete boundary of a portion of the given region. A region that does not satisfy the definition of

* See Goursat-Hedrick, *Mathematical Analysis*, Art. 97.

a simply connected region is called a **multiply connected region.** The region S shown in Fig. 18 is a multiply connected region, since it is possible to have a closed curve surrounding one of the circles that does not of itself form a complete boundary of a portion of the region

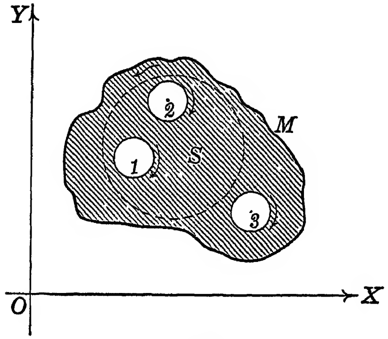

Fig. 18.

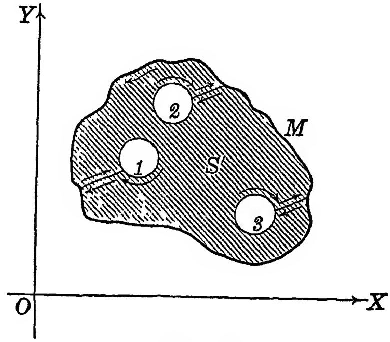

Fig. 19.

inclosed. In a simply connected region, a curve between two points may always be changed by continuous deformation into any other curve between those points, both curves lying completely in the region considered; this is not true for a multiply connected region.

A finite, multiply connected region may be changed into a simply connected region by drawing from each inner portion of the boundary a line connecting it with the outer portion. A line joining two points of the boundary is called a **cross-cut,** and we shall so choose the line that it will neither intersect itself nor other cross-cuts. For example, the multiply connected region S shown in Fig. 18 may be changed into a simply connected region by connecting the circles 1, 2, 3 with the boundary curve M by cross-cuts as indicated in Fig. 19. It serves the purpose equally well to connect each of the circles with some one other circle by means of cross-cuts and one of them with the outer boundary M. Thus in Fig. 19 we might have joined circle 2 to circle 3 and to circle 1 and then have joined any one of the three circles to the contour M. The boundary points connected by a cross-cut may be distinct, as in the case of the cross-cuts joining the separate curves constituting a portion of the boundary in Fig. 19 or the cross-cut a, b in Fig. 20; or we may have the two boundary points coincident as in the case of the cross-cut drawn from the point c and returning to the same point, in which case the cross-cut is a closed curve.

The notion of a path of integration may now be extended so as to include the boundary of a multiply connected region; for, having made the region simply connected by the introduction of cross-cuts, the integral may be taken along the contour of this simply connected region including the cross-cuts. For example, in Fig. 19 the arrows indicate the positive direction of the integral with respect to the region S'. It will be noted that in any such case the integral is taken twice along each of the cross-cuts, once in either direction. This portion of the integral vanishes in accordance with the second property of line-integrals already stated. We may then say that the integral taken over the contour of a multiply connected region, that is, the sum of the integrals taken over the several closed curves constituting the boundary, with the proper signs attached, is the same as the integral taken over the boundary of the simply connected region formed by inserting cross-cuts in the given region.*

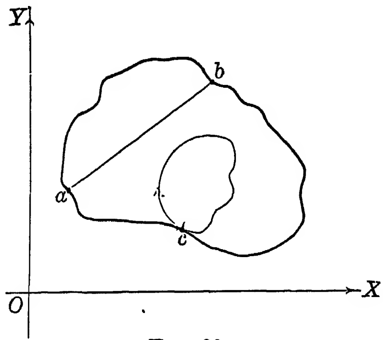

FIG. 20.

16. Green's theorem. One of the important theorems associated with line-integrals gives, under certain conditions, a relation between such integrals and ordinary double integrals. This theorem, known as **Green's theorem,** may be stated for functions of two real variables as follows:

THEOREM I. *In a given finite region S let C be the complete boundary of any portion of the plane such that C lies within S and incloses only points of S. If in the given region $P(x, y)$ and $Q(x, y)$ are continuous real functions of x and y together, having the continuous partial derivatives* $\frac{\partial Q}{\partial x}$, $\frac{\partial P}{\partial y}$, *then*

$$\int_C P\,dx + Q\,dy = \int\int\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dx\,dy,$$

where the double integral is to be taken over the region bounded by C.

* Cf. Osgood, *Lehrbuch der Funktionentheorie*, 2d Ed., Vol. I, Chap. IV, Art. 4, Chap. V, Art. 7.

Let us consider the double integral

$$\int\int \frac{\partial P}{\partial y}\, dx\, dy,$$

taken over the region bounded by the curve C. We shall first consider this curve to be a single closed curve such that any parallel to the Y-axis cuts it in at most two points as indicated in Fig. 21. Let A and B be the points of C at which x has a minimum and a maximum, respectively. A parallel to the Y-axis lying between Aa and Bb cuts the curve C in two points whose ordinates may be denoted by y_1, y_2, respectively.

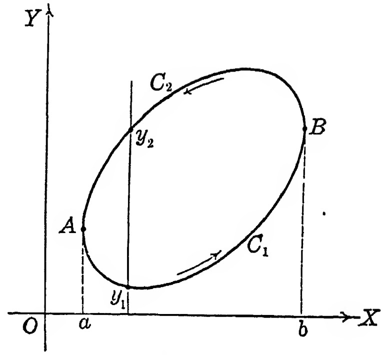

FIG. 21.

Because of the continuity of $\frac{\partial P}{\partial y}$, the following integrals exist, and we may write *

$$\int\int \frac{\partial P}{\partial y}\, dx\, dy = \int_a^b dx \int_{y_1}^{y_2} \frac{\partial P}{\partial y}\, dy. \tag{1}$$

Performing the integration in the second member with respect to y, we get from this equation the following relation:

$$\int\int \frac{\partial P}{\partial y}\, dx\, dy = \int_a^b \{P(x, y_2) - P(x, y_1)\}\, dx. \tag{2}$$

However, the two integrals

$$\int_a^b P(x, y_2)\, dx, \qquad \int_a^b P(x, y_1)\, dx$$

are line-integrals taken along the paths Ay_2B, Ay_1B, respectively. Hence, instead of the integral in the second member of (2) we may write a single integral taken in a negative direction around the curve C, or what is the same, we may write the negative of that integral taken in a positive direction around C, and thus have

$$\int\int \frac{\partial P}{\partial y}\, dx\, dy = -\int_C P\, dx. \tag{3}$$

* See Hobson, *Theory of Functions of a Real Variable*, Arts. 314 and 315.

In a similar manner, we can deduce the relation

$$\int\int \frac{\partial Q}{\partial x}\, dx\, dy = \int_C Q\, dy. \tag{4}$$

By subtracting (3) from (4) we have

$$\int_C P\, dx + Q\, dy = \int\int \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy,$$

as the theorem requires.

To extend (3) and (4) to any finite region bounded by an ordinary curve C, all that is needed is to divide the region into subregions, each of which satisfies the condition that a line parallel to the Y-axis, in case of (3), or to the X-axis, in case of (4), cuts the boundary curve in not more than two points, or in a segment as p, q, Fig. 22. This is always possible as indicated in Fig. 22, since the given contour by hypothesis has only a finite number of maxima and minima with respect to x and with respect to y. If the boundary C consists of several closed curves, the region is multiply connected. By the proper introduction of cross-cuts this region may be made simply connected, and the foregoing argument then holds. As we have seen, however, the value of the integral along C becomes in this case the sum of the integrals taken in the proper direction along the several closed curves composing C.

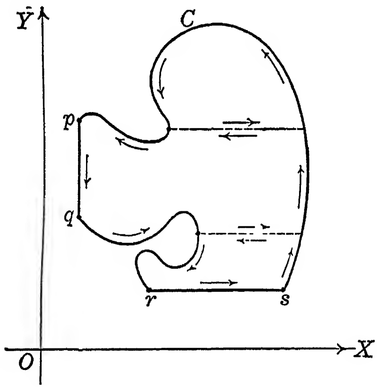

Fig. 22.

We shall now consider the following theorem, which is important in subsequent discussions.

Theorem II. *In a given finite region S let C be the complete boundary of any portion of the plane such that C lies within S and incloses only points of S. If in the given region $P(x, y)$, $Q(x, y)$ are continuous real functions of x and y together, having the continuous partial derivatives $\frac{\partial Q}{\partial x}$, $\frac{\partial P}{\partial y}$, then the necessary and sufficient condition that the integral*

$$\int P\, dx + Q\, dy \tag{5}$$

vanishes for every such curve C is that

$$\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y} \tag{6}$$

for all points of S.

That this condition is necessary may be established as follows. We have given the condition that the line-integral (5) vanishes to show that the condition (6) follows as a consequence. The function $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ is continuous in S. If this function is not identically zero for all values of x, y in S, then it is possible to find a subregion R sufficiently small such that $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ is of the same sign for all values of x, y in R. From Green's theorem, we have

$$\int_C P\,dx + Q\,dy = \int\!\!\int \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy. \tag{7}$$

If in R the function $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ is always of the same sign for all values of x, y, then the double integral in the second member of (7) can not vanish and consequently the line-integral $\int_C P\,dx + Q\,dy$ can not equal zero when taken around the contour of R. Hence, in order that the integral (5) taken around every complete boundary C in S shall vanish, the condition (6) must be satisfied identically for all values of x, y in S.

That the condition stated in the theorem is also sufficient follows at once from equation (7); for, if (6) holds for all values of x, y in S, then the integral (5) taken along any complete boundary C in S vanishes as the theorem requires.

We are now in position to establish the following proposition.

THEOREM III. *In a given finite simply connected region S, let L be any ordinary curve joining two points of S and lying within S. If in the given region P(x, y), Q(x, y) are continuous real functions with respect to x and y together, having the continuous partial derivatives* $\frac{\partial Q}{\partial x}$, $\frac{\partial P}{\partial y}$, *then the necessary and sufficient condition that the line-integral* $\int_L P\,dx + Q\,dy$ *is independent of the path L is that*

$$\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y}$$

for all points of S.

This theorem follows directly from the conclusions of Theo... Let (x_0, y_0), (x_1, y_1) be any two points in the region S (Fig. 23). Let L_1, L_2 be any two non-intersecting ordinary curves joining these points and lying wholly in S. The lines L_1, L_2 taken together constitute a closed curve C lying in S; and if $\dfrac{\partial Q}{\partial x} = \dfrac{\partial P}{\partial y}$ for all points in S, then by Theorem II the integral $\int_C P\,dx + Q\,dy$ is zero. Hence the two integrals

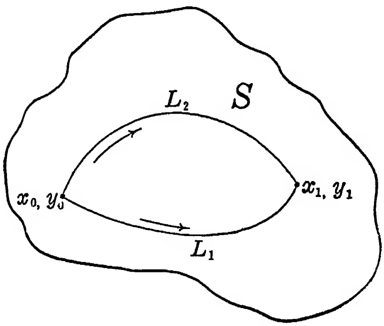

FIG. 23.

$$\int_{L_1} P\,dx + Q\,dy, \quad \int_{L_2} P\,dx + Q\,dy$$

must be equal, both integrals being taken in a positive direction from (x_0, y_0) to (x_1, y_1). In other words, the line-integral is independent of the path. Conversely, if these two integrals are equal, the integral around the closed curve C vanishes; but C is any ordinary closed curve in S and hence by Theorem II, we have

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x},$$

as the theorem requires.

To apply the foregoing theorem to a multiply connected region, it is necessary first to make it simply connected by the introduction of the proper cross-cuts.

The integral $\displaystyle\int_{x_0, y_0}^{x, y} P\,dx + Q\,dy$ taken along an arbitrary path starting from a fixed point (x_0, y_0) and having a variable upper limit is, under the conditions set forth in Theorem III, a function of x and y. Moreover, we have the following theorem.

THEOREM IV. *In a given finite simply connected region S, let (x_0, y_0) be any fixed point and (x, y) a variable point. If in the given region the functions $P(x, y)$, $Q(x, y)$ are continuous real functions in both x and y having the continuous partial derivatives $\dfrac{\partial Q}{\partial x}, \dfrac{\partial P}{\partial y}$, satisfying the condition $\dfrac{\partial Q}{\partial x} = \dfrac{\partial P}{\partial y}$, then the integral*

$$\int_{x_0, y_0}^{x, y} P\,dx + Q\,dy$$

defines a function $F(x, y)$ such that

$$\frac{\partial F}{\partial x} = P, \qquad \frac{\partial F}{\partial y} = Q.$$

From what has been said, we know that the given integral defines some function of the variable upper limit. We designate this function by $F(x, y)$ and proceed to show that this function satisfies the condition of the theorem. Let (x_1, y_1) be any point of S and let $(x_1 + \Delta x, y_1)$ be a second point of S in the neighborhood of (x_1, y_1). We have then

$$\begin{aligned} F(x_1 + \Delta x, y_1) - F(x_1, y_1) &= \int_{x_0,\, y_0}^{x_1+\Delta x,\, y_1} P\,dx + Q\,dy - \int_{x_0,\, y_0}^{x_1,\, y_1} P\,dx + Q\,dy \\ &= \int_{x_1,\, y_1}^{x_1+\Delta x,\, y_1} P\,dx + Q\,dy. \end{aligned}$$

Since by Theorem III the path of integration is arbitrary, we may assume it to be rectilinear, and hence we can write

$$\int_{x_1,\, y_1}^{x_1+\Delta x,\, y_1} P\,dx + Q\,dy = \int_{x_1}^{x_1+\Delta x} P\,dx; \tag{8}$$

for, as y does not vary the value of dy is zero. The resulting integral being an ordinary definite integral, we can apply the first theorem of the mean * for such integrals and thus obtain

$$\int_{x_1}^{x_1+\Delta x} P\,dx = P(x_1 + \theta\Delta x, y_1)\,\Delta x, \qquad 0 < \theta < 1. \tag{9}$$

From (8) and (9) we now obtain

$$\frac{F(x_1 + \Delta x, y_1) - F(x_1, y_1)}{\Delta x} = P(x_1 + \theta\Delta x, y_1).$$

Taking the limit as $\Delta x \doteq 0$, we have

$$\left.\frac{\partial F}{\partial x}\right]_{x_1,\, y_1} = P(x_1, y_1).$$

In a similar manner, we may show that

$$\left.\frac{\partial F}{\partial y}\right]_{x_1,\, y_1} = Q(x_1, y_1).$$

Since x_1, y_1 is any point of S, we may write

$$\frac{\partial F}{\partial x} = P(x, y), \qquad \frac{\partial F}{\partial y} = Q(x, y),$$

for all values of x, y in S.

* See Goursat-Hedrick, *Mathematical Analysis*, p. 151.

17. Integral of $f(z)$. We shall now consider the integral of a function of a complex variable. Let $f(z)$ be a continuous single-valued function of z along a given ordinary curve C joining the two points α, β. Between the points $\alpha \equiv z_0$ and $\beta \equiv z_n$ insert $n - 1$ division points of the curve, $z_1, z_2, \ldots, z_{n-1}$. Form the sum of the products $f(\zeta_k)\Delta_k z$, where $\Delta_k z$ is the difference $z_k - z_{k-1}$ and $f(\zeta_k)$ is the value of the given function $f(z)$ at some point ζ_k on the curve C between z_{k-1} and z_k. The integral of $f(z)$ along C between the limits α and β is defined as the limit of this sum as the number of division points between α and β is increased indefinitely and each difference $| z_k - z_{k-1} |$ approaches zero; that is,

$$\int_C f(z)\,dz = \underset{n=\infty}{L} \sum_{k=1}^{n} f(\zeta_k)\,\Delta_k z. \tag{1}$$

The existence of this integral can be established either directly, as in the case of real variables, or by making it depend upon the existence of the line-integrals already discussed. We shall choose the latter method. For this purpose we write

$$f(z) = u(x, y) + iv(x, y),$$

where u, v are real functions of the real variables x, y, and hence we have by putting $\zeta_k = \xi_k + i\eta_k$,

$$\begin{aligned}\sum_{k=1}^{n} f(\zeta_k)\,\Delta_k(z) &= \sum_{k=1}^{n} f(\zeta_k)\,(z_k - z_{k-1}) \\ &= \sum_{k=1}^{n} \{u(\xi_k, \eta_k)(x_k - x_{k-1}) - v\,(\xi_k, \eta_k)(y_k - y_{k-1})\} \\ &\quad + i\sum_{k=1}^{n} \{v(\xi_k, \eta_k)(x_k - x_{k-1}) + u(\xi_k, \eta_k)(y_k - y_{k-1})\}.\end{aligned}$$

Since $f(z)$ is continuous in z along C, it follows from Theorem II, Art. 13, that both $u(x, y)$, $v(x, y)$ are continuous in x, y together along the same curve. Moreover, as $\Delta z \doteq 0$, we have $\Delta x \doteq 0$, $\Delta y \doteq 0$. Hence, upon passing to the limit, we obtain

$$\int_C f(z)\,dz = \int_C (u\,dx - v\,dy) + i\int_C (v\,dx + u\,dy); \tag{2}$$

for, from the discussion of line-integrals it follows that the two integrals in the second member of this equation both exist because of

the continuity of $u(x, y)$, $v(x, y)$, and hence the integral $\int_C f(z)\,dz$ must also exist.

In the case of simple functions the integral between two given points on a given curve may be evaluated by direct application of the definition (1).

Ex. 1. Evaluate the integral $\int_\alpha^\beta dz$.

This integral is independent of the path over which it is taken; for, by definition, we have

$$\int_\alpha^\beta dz = \underset{n=\infty}{L} \sum_{k=1}^{n} (z_k - z_{k-1}) = \underset{n=\infty}{L} (z_1 - z_0 + z_2 - z_1 + z_3 - z_2 + \cdots + z_n - z_{n-1})$$
$$= \underset{n=\infty}{L} (z_n - z_0) = \beta - \alpha,$$

since $z_0 \equiv \alpha$, $z_n \equiv \beta$, no matter what the intermediate points may be.

The geometric interpretation (Fig. 24) of this result will enable the student to understand more clearly the nature of a definite integral of a function of a complex variable. Such an integral was defined as the limit of the sum $\sum f(\zeta_k)(z_k - z_{k-1})$; that is, we consider the limit of a sum of products, each of which is the value of the given function at some point ζ_k on the curve multiplied into the directed chord $z_k - z_{k-1}$. In this particular case, the value of $f(\zeta_k)$ is always unity. Adding $z_1 - z_0$ to $z_2 - z_1$, we have geometrically the directed chord $z_2 - z_0$. Adding to this result the directed chord $z_3 - z_2$, we have the directed chord $z_3 - z_0$, etc. Finally, we have $z_n - z_0$, which is identically $\beta - \alpha$ as we have seen. This result is very different from taking the integral $\int_\alpha^\beta |\,dz\,|$. In this case, we add merely the chords without reference to direction; that is, we have $\underset{n=\infty}{L} \sum_{k=1}^{n} |\,\Delta_k z\,|$. In the limit we should have in this case not $\beta - \alpha$ but the length L of the path * of integration from α to β.

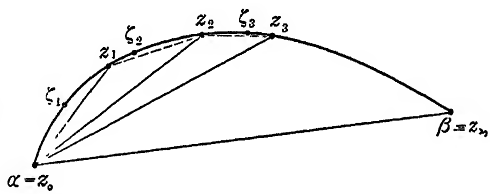

FIG. 24.

* See Townsend and Goodenough, *First Course in Calculus*, Art. 88.

Ex. 2. Evaluate $\int_\alpha^\beta z\,dz$.

This integral is independent of the path over which it is taken; for, the limit defining the integral exists when ζ_k is any point in the interval $z_{k-1}, - z_k$. We may, therefore, select for ζ_k any convenient point in this interval. If we take it to be z_k or z_{k-1}, we have respectively

$$\int_\alpha^\beta z\,dz = \underset{n=\infty}{L} \sum_{k=1}^{n} z_k(z_k - z_{k-1}), \quad \text{or} \quad \int_\alpha^\beta z\,dz = \underset{n=\infty}{L} \sum_{k=1}^{n} z_{k-1}(z_k - z_{k-1}).$$

Hence, we have by taking one-half of the sum of these two results

$$\int_\alpha^\beta z\,dz = \frac{\underset{n=\infty}{L} \sum_{k=1}^{n} (z_k^2 - z_{k-1}^2)}{2}$$

$$= \underset{n=\infty}{L} \frac{(z_1^2 - z_0^2 + z_2^2 - z_1^2 + z_3^2 - z_2^2 + \cdots + z_n^2 - z_{n-1}^2)}{2}$$

$$= \frac{\underset{n=\infty}{L} (z_n^2 - z_0^2)}{2} = \frac{(\beta^2 - \alpha^2)}{2},$$

no matter what the path is, since as in Ex. 1, $z_0 \equiv \alpha$, $z_n \equiv \beta$.

In both of these examples the result obtained is the same as that obtained by substituting the limits of integration in a function of which the integrand is the derivative and taking the difference.

The definite integral of a function of a complex variable has been defined as a limit. From this definition and the laws of operation with limits, the general properties of such integrals can be deduced; or, they may be shown to hold as a consequence of the corresponding properties of line-integrals in the calculus of real variables. The proof in many cases is so evident that it is left to the reader to supply. If α, β be two points on the path of integration C, we then have among other properties:

1. $\int_\beta^\alpha f(z)\,dz = -\int_\alpha^\beta f(z)\,dz.$

2. $\int_\alpha^\beta cf(z)\,dz = c\int_\alpha^\beta f(z)\,dz.$

3. $\int_\alpha^\beta [f(z) \pm \phi(z)]\,dz = \int_\alpha^\beta f(z)\,dz \pm \int_\alpha^\beta \phi(z)\,dz.$

This last property can be readily extended to the case involving any finite number of functions. It can not, however, be extended

to the case involving an infinite number of functions without introducing some condition as to the character of the convergence of the series thus introduced.

4. $\displaystyle\int_\alpha^\beta f(z)\,dz = \int_\alpha^{z_1} f(z)\,dz + \int_{z_1}^\beta f(z)\,dz,$

where z_1 lies upon the path of integration C connecting α and β.

This property can be extended to the case where the path of integration is broken up into a finite number of parts by inserting between α and β, $n-1$ points $z_1, z_2, \ldots, z_{n-1}$ on the path of integration. If the ordinary curve constituting the path of integration C is composed of a finite number of connected lines C_1, $C_2, \ldots, C_n$, we write

$$\int_C f(z)\,dz = \int_{C_1} f(z)\,dz + \int_{C_2} f(z)\,dz + \cdots + \int_{C_n} f(z)\,dz.$$

This relation enables us to extend the definition of an integral to include integrals taken over the contour of a multiply connected region. As in line-integrals of functions of real variables, the integral is said to be taken in a **positive direction** with respect to the region bounded when it is taken in a positive direction with respect to this region about each closed curve constituting a portion of the boundary. The integral over the complete boundary is unaffected by the introduction of the cross-cuts necessary to make the given region simply connected.

5. $$\left|\int_\alpha^\beta f(z)\,dz\right| \leqq \int_\alpha^\beta |f(z)|\cdot|dz|.$$

We have

$$\left|\sum_{k=1}^n f(\zeta_k)\,\Delta_k z\right| \leqq \sum_{k=1}^n |f(\zeta_k)\,\Delta_k z| = \sum_{k=1}^n |f(\zeta_k)|\cdot|\Delta_k z|,$$

since the absolute value of a sum is less than or at most equal to the sum of the absolute values of the terms, and the absolute value of a product is always equal to the product of the absolute values of the factors. Passing to the limit as Δz approaches zero, we have the required relation.

From Ex. 1, we have

6. $$\int_\alpha^\beta |dz| = L,$$

where L is the length of the path of integration.

7. $$\left| \int_\alpha^\beta f(z)\, dz \right| \leqq M \cdot L,$$

where M is the maximum value of $|f(z)|$ along the path of integration and L is the length of that path.

The result stated in this theorem follows at once from (5) and (6); for, we have upon replacing $|f(z)|$ by its maximum value

$$\left| \int_\alpha^\beta f(z)\, dz \right| \leqq M \int_\alpha^\beta |dz| = M \cdot L.$$

18. Change from complex to real variable. We can readily deduce the law for the change of the independent variable in a definite integral of a complex variable. We shall first consider the change from a complex variable to a real variable. We have from equation (2), Art. 17,

$$\int_{AB} f(z)\, dz = \int_{AB} u\, dx - v\, dy + i \int_{AB} v\, dx + u\, dy, \tag{1}$$

where $$f(z) = u(x, y) + iv(x, y).$$

By aid of equation (11), Art. 15, we may express the two integrals in the second member of (1) in terms of a parameter t and thus obtain

$$\int_{AB} u\, dx - v\, dy = \int_{t_0}^{t_n} \{u \cdot \Psi_1'(t) - v \cdot \Psi_2'(t)\}\, dt, \tag{2}$$

$$\int_{AB} v\, dx + u\, dy = \int_{t_0}^{t_n} \{v \cdot \Psi_1'(t) + u \cdot \Psi_2'(t)\}\, dt, \tag{3}$$

where as in Art. 15

$$x = \Psi_1(t), \qquad y = \Psi_2(t), \qquad t_0 \leqq t \leqq t_n$$

are the parametric equations of the path of integration. By combining (2) and (3), we have from (1),

$$\begin{aligned} \int_{AB} f(z)\, dz &= \int_{t_0}^{t_n} \{u \cdot \Psi_1'(t) - v \cdot \Psi_2'(t)\}\, dt + i \int_{t_0}^{t_n} \{v \cdot \Psi_1'(t) + u \cdot \Psi_2'(t)\}\, dt \\ &= \int_{t_0}^{t_n} (u + iv)\, \{\Psi_1'(t) + i\Psi_2'(t)\}\, dt. \end{aligned} \tag{4}$$

Remembering that

$$D_t z = D_t[\Psi_1(t) + i\Psi_2(t)] = \Psi_1'(t) + i\Psi_2'(t),$$

and putting

$$\begin{aligned} f(z) &= u(x, y) + iv(x, y) \\ &= u\,\{\Psi_1(t), \Psi_2(t)\} + iv\{\Psi_1(t), \Psi_2(t)\} \\ &= F(t), \end{aligned}$$

we may write the above result in the following compact form:

$$\int_{AB} f(z)\,dz = \int_{t_0}^{t_n} F(t)\,D_t z\,dt. \tag{5}$$

We thus obtain precisely the same rule for change of variable in the integral $\int_{AB} f(z)\,dz$ as is ordinarily formulated for the definite integral of a function of a real variable; namely, substitute in $f(z)\,dz$

$$z = \Psi_1(t) + i\Psi_2(t), \qquad dz = \{\Psi_1'(t) + i\Psi_2'(t)\}\,dt,$$

with a corresponding change in the limits and the path connecting them. Equation (4), or its equivalent equation (5), gives a means by which the calculation of an integral of a function of a complex variable may be made to depend upon the evaluation of at most four ordinary definite integrals of functions of a real variable.

Ex. 1. Evaluate the integral $\int_C \frac{dz}{z-\alpha}$, where C is a circle of radius ρ about α.

Put
$$z - \alpha = \rho(\cos\theta + i\sin\theta),$$
whence
$$\begin{aligned} D_\theta z &= \rho(-\sin\theta + i\cos\theta) \\ &= i\rho(\cos\theta + i\sin\theta). \end{aligned}$$

As z describes the circle C, θ passes from 0 to 2π. We have then

$$\begin{aligned} \int_C \frac{dz}{z-\alpha} &= \int_0^{2\pi} \frac{1}{\rho(\cos\theta + i\sin\theta)} \cdot i\rho(\cos\theta + i\sin\theta)\,d\theta \\ &= \int_0^{2\pi} i\,d\theta \\ &= 2\pi i. \end{aligned}$$

Ex. 2. Evaluate the integral $\int_C \frac{dz}{(z-\alpha)^n}$, where C is the same as in Ex. 1, and n is an integer different from one.

As in the preceding example, put

$$z - \alpha = \rho(\cos\theta + i\sin\theta).$$

We then have

$$\begin{aligned} \int_C \frac{dz}{(z-\alpha)^n} &= \int_0^{2\pi} \frac{i\rho(\cos\theta + i\sin\theta)\,d\theta}{\rho^n(\cos\theta + i\sin\theta)^n} \\ &= \frac{i}{\rho^{n-1}} \int_0^{2\pi} \{\cos(n-1)\,\theta - i\sin(n-1)\,\theta\}\,d\theta \\ &= \frac{i}{\rho^{n-1}} \left\{ \int_0^{2\pi} \cos(n-1)\,\theta\,d\theta - i\int_0^{2\pi} \sin(n-1)\,\theta\,d\theta \right\}. \end{aligned}$$

But from the calculus of real variables, we have for $n \neq 1$,

$$\int_0^{2\pi} \cos (n-1)\,\theta\, d\theta = 0, \qquad \int_0^{2\pi} \sin (n-1)\,\theta\, d\theta = 0.$$

Hence, we obtain the result

$$\int_C \frac{dz}{(z-\alpha)^n} = 0, \qquad n \neq 1.$$

We shall consider later (Art. 22) the case of change of variable from one complex variable to another.

19. Cauchy-Goursat theorem. The properties of definite integrals considered in Art. 17 depend upon the condition that $f(z)$ is a continuous function. Consequently, they hold where the given function is holomorphic, since such a function is necessarily continuous.

We shall now consider some of the special properties of integrals of functions which are holomorphic in a given region. The most important and fundamental of these properties is stated in a theorem due originally to Cauchy. The proof of this theorem may be made to depend upon Green's theorem but the results obtained can then be said to hold only under the initial restrictions assumed in the demonstration of that theorem. Goursat has shown * that the condition that the derived function $f'(z)$ is continuous is not necessary for the demonstration. In order to establish the theorem without assuming this condition, the following lemma will be of use.

LEMMA. *Given a region T in which $f(z)$ is holomorphic. Let the ordinary closed curve C, lying wholly in T and containing only points of T, be the complete boundary of a region T'. It is always possible to divide the region T' into a finite number of squares S_i and partial squares R_i such that within or upon the boundary of each of these subregions there exists a point z_i such that as z describes the boundary of the subregion we have*

$$\left| \frac{f(z) - f(z_i)}{z - z_i} - f'(z_i) \right| < \epsilon, \tag{1}$$

where ϵ is a previously assigned arbitrarily small positive number.

The boundary curve C (Fig. 25) is by hypothesis an ordinary curve and hence has but a finite number of maxima and minima with respect to x and with respect to y. Consequently, by drawing lines parallel to

* See *Trans. Amer. Math. Soc.*, Vol. I, pp. 14–16; Moore, *Ibid.*, pp. 499–506; Pringsheim, *Trans. Amer. Math. Soc.*, Vol. II, pp. 413–421.

the two axes of coördinates we can cover the given region T' with a system of congruent squares, such that the perimeter of no square is cut by the contour C in more than two points. Let c denote the length of a side of these squares and let A denote the combined area of these squares. By this means the given region T' is subdivided into smaller regions. Some of these regions are complete squares and others are partial squares along the contour C, bounded in part by straight line segments and in part by arcs of C. There may or may not exist within or upon the boundary of each of these subregions a point z_1 satisfying the conditions of the lemma.

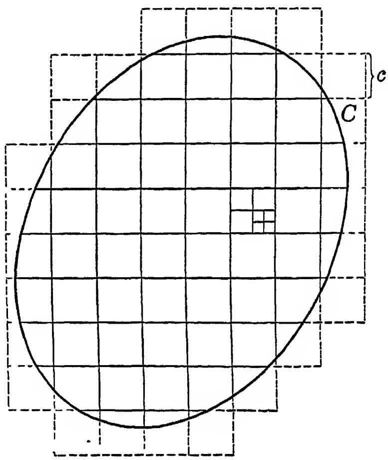

FIG. 25.

If in any subregion such a point does not exist, we divide the corresponding square into four equal squares by drawing lines parallel to the two axes of coördinates. Those squares or partial squares satisfying the condition given in (1) are left unchanged. Moreover, we consider only those new squares and partial squares that lie in the region T'. If any of these new squares or parts of squares satisfy the required condition, they are not subjected to further subdivision. The process of subdivision is, however, continued with the rest. In this way either there is ultimately obtained a finite number of subregions of the desired character, or there exists at least one infinite sequence of squares each lying within the preceding, such that these squares or, in case the squares contain points not in T', the corresponding partial squares in no case satisfy the condition set forth in (1). The sides of the squares of this infinite sequence approach zero as a limit, and the sequence satisfies the conditions of Theorem IV of Art. 12. Consequently, such a sequence defines a definite limiting point α.

The point α is a regular point of the function $f(z)$; hence the derivative $f'(\alpha)$ exists and we have

$$\underset{z \doteq \alpha}{L} \frac{f(z)-f(\alpha)}{z-\alpha} = f'(\alpha), \tag{2}$$

where $z - \alpha$ is the increment of z. This relation can be written in the form

$$\left| \frac{f(z) - f(\alpha)}{z - \alpha} - f'(\alpha) \right| < \epsilon, \tag{3}$$

which holds for all values of z in T' such that $|z - \alpha| < \delta$. Draw about the point α a circle of radius δ. From some point on in the sequence of regions defining the point α, all of the regions lie within this circle, and consequently if α is taken as the point z_i the values of z upon the perimeter of any one of these regions are such that (1) is satisfied. This conclusion contradicts the assumption that none of the subregions of the sequence satisfies the required condition. From this contradiction the given proposition follows.

By aid of this lemma, we may now demonstrate the **Cauchy-Goursat theorem,** which may be stated as follows:

THEOREM I. *Let $f(z)$ be holomorphic in a given finite region S and let C be the complete boundary of any portion S' of S such that C lies wholly in S and incloses only points of S; then*

$$\int_C f(z)\, dz = 0.$$

The boundary C may consist of a single closed ordinary curve or a combination of such curves. We shall first consider the case where C is a single closed ordinary curve and the inclosed region S' is simply connected. Let S' be divided into squares and partial squares satisfying condition (1) of the foregoing lemma. Let n denote the number of squares S_i and m the number of partial squares R_i. If the integral is taken in a positive direction around the perimeter of the various subregions S_i, R_i, it will be seen that each side of these regions that is not a portion of C is taken twice as a path of integration, the two integrals being taken however in opposite direction. Considering the sum of the integrals about the perimeters of all of the regions S_i, R_i, we may therefore write by aid of 4, Art. 17,

$$\int_C f(z)\, dz = \sum_{i=1}^{n} \int_{\gamma_i} f(z)\, dz + \sum_{i=1}^{m} \int_{\lambda_i} f(z)\, dz, \tag{4}$$

where γ_i, λ_i denote the boundaries of S_i, R_i, respectively.

From the lemma, we have within or upon the boundary of each S_i, R_i a point z_i such that

$$\left| \frac{f(z) - f(z_i)}{z - z_i} - f'(z_i) \right| < \epsilon. \tag{5}$$

This relation can be written in the form

$$f(z) - f(z_\iota) = (z - z_\iota) f'(z_\iota) + \eta_\iota (z - z_\iota), \tag{6}$$

where η_ι is a function of z such that

$$|\eta_\iota| < \epsilon, \tag{7}$$

when z varies along the contour of S_ι or R_ι. We shall now consider the integral around the perimeter of one of the squares S_ι. We have from (6), since z_ι is a constant for this integration,

$$\int_{\gamma_\iota} f(z)\,dz = [f(z_\iota) - z_\iota f'(z_\iota)] \int_{\gamma_\iota} dz + f'(z_\iota) \int_{\gamma_\iota} z\,dz + \int_{\gamma_\iota} \eta_\iota (z - z_\iota)\,dz. \tag{8}$$

From Exs. 1, 2, Art. 17, we know that

$$\int_\alpha^\beta dz = \beta - \alpha, \qquad \int_\alpha^\beta z\,dz = \tfrac{1}{2}(\beta^2 - \alpha^2).$$

In the particular case under consideration, as the path of integration is a closed curve, α and β are the same point, and hence both of these integrals vanish. From equation (8) we then have

$$\left| \int_{\gamma_\iota} f(z)\,dz \right| = \left| \int_{\gamma_\iota} \eta_\iota (z - z_\iota)\,dz \right| \cdot \tag{9}$$

Let the length of one side of the square S_ι be c_ι. The diagonal of the square is then $c_\iota \sqrt{2}$. Hence, we have

$$|z - z_\iota| \leqq c_\iota \sqrt{2}.$$

Making use of this relation and of that given in (7), we may now write by 7, Art. 17,

$$\left| \int_{\gamma_\iota} f(z)\,dz \right| < \epsilon\, c_\iota \sqrt{2} \int_{\gamma_\iota} |dz| = \epsilon\, c_\iota \sqrt{2} \cdot 4\, c_\iota = \epsilon\, 4\sqrt{2} A_\iota, \tag{10}$$

where A_ι denotes the area of S_ι.

Consider now the integral taken around one of the partial squares R_ι. We have

$$\int_{\lambda_\iota} f(z)\,dz = [f(z_\iota) - z_\iota f'(z_\iota)] \int_{\lambda_\iota} dz + f'(z_\iota) \int_{\lambda_\iota} z\,dz + \int_{\lambda_\iota} \eta_\iota (z - z_\iota)\,dz. \tag{11}$$

As before the first two integrals in the second member of this equation vanish and we have

$$\left| \int_{\lambda_\iota} f(z)\,dz \right| = \left| \int_{\lambda_\iota} \eta_\iota (z - z_\iota)\,dz \right| \cdot \tag{12}$$

We may denote by c_i the length of a side of the square of which R_i is a portion, R_i being that portion of the square cut off by curve C and lying in S'. Let l_i be the length of that arc of C which forms a portion of the boundary of R_i. We have then $|z - z_i| \leqq c_i\sqrt{2}$. From (12), we have

$$\left|\int_{\lambda_i} f(z)\,dz\right| < \epsilon c_i\sqrt{2}\int_{\lambda_i}|dz| < \epsilon c_i\sqrt{2}(4c_i+l_i) \leqq \epsilon\sqrt{2}(4B_i+cl_i), \quad (13)$$

where B_i denotes the area of the square of which R_i is a part, and c is the length of one side of the largest square that comes into consideration in the subdivision of S'.

Replacing each term of the sums in (4) by its absolute value, we have, by use of (10) and (13),

$$\left|\int_C f(z)\,dz\right| < \epsilon\sqrt{2}\left\{\sum_{i=1}^{n} 4A_i + \sum_{i=1}^{m}(4B_i + cl_i)\right\}$$
$$\leqq \epsilon\sqrt{2}\{4A + cL\}, \quad (14)$$

where L denotes the length of the curve C, and A denotes, as in the discussion of the lemma, the combined area of the system of congruent squares with which the region S' was originally covered. The expression included in the braces is therefore a constant, and as ϵ is arbitrarily small the product is arbitrarily small. As the absolute value of the integral $\int_C f(z)\,dz$ is shown to be less than an arbitrarily small number, it follows that

$$\int_C f(z)\,dz = 0,$$

as required by the theorem.

The above theorem is now established for the case where the region S', bounded by C, is a simply connected region. The proof may readily be extended to the case where S' is multiply connected. By introducing the necessary cross-cuts, S' becomes simply connected and the foregoing proof applies. However, in taking the integral around the boundary, including the cross-cuts, these cross-cuts are traversed twice, once in each direction. As pointed out in Art. 17, this portion of the integral vanishes, and we have the theorem applying to the complete boundary C of the multiply connected region S'.

In the statement of Theorem I, it is assumed that $f(z)$ is holomorphic in the region S' including its boundary C. We shall now show that it is sufficient that $f(z)$ is holomorphic within the region bounded by C and converges uniformly to its values along C. As we have already seen (Art. 13), such a condition is equivalent to saying that the values of $f(z)$ along C are continuous with the values of the function within the region bounded by C. Moreover, uniform convergence, together with continuity of $f(z)$ within the region bounded, enables us to say that $f(z)$ changes continuously as z varies continuously along C. In certain discussions, the conditions given in the following theorem will be more convenient than those of Theorem I.

THEOREM II. *If $f(z)$ is holomorphic within a finite region S bounded by an ordinary curve C and if it converges uniformly to its values along C, then*

$$\int_C f(z)\,dz = 0.$$

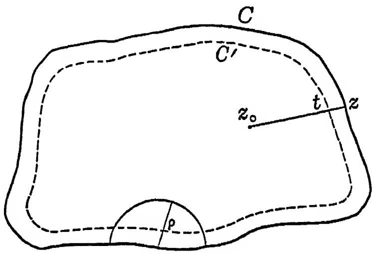

FIG. 26.

Since $f(z)$ converges uniformly along C, it follows from the discussion of uniform convergence (Art. 13) that about each point z of C there may be drawn a partial circle of radius ρ, which is the independent of z, such that for all values of t within this partial circle we have

$$|f(t) - f(z)| < \epsilon, \qquad |t - z| < \rho.$$

Since this condition holds simultaneously for all values of z along C, there exists a closed curve C' such that as t traverses C' we have

$$t = z_0 + \theta(z - z_0), \qquad 0 < \theta < 1,$$

where z_0 is a fixed point interior to the region bounded by C' and θ is a constant. The difference of the integrals of the given function taken along the two curves C and C' gives*

$$\int_C f(z)\,dz - \int_{C'} f(t)\,dt = \int_C f(z)\,dz - \theta\int_C f[z_0 + \theta(z - z_0)]\,dz$$
$$= \int_C \{f(z) - \theta f[z_0 + \theta(z - z_0)]\}\,dz. \qquad (15)$$

* See Art. 22 for method of change of variable.

But the integrand in the last of these integrals may be written in the form

$$\begin{aligned} f(z) - \theta f[z - (z - z_0)(1-\theta)] \\ = f(z) - f[z-(z-z_0)(1-\theta)] + (1-\theta) f[z-(z-z_0)(1-\theta)]. \end{aligned}$$

Since $(1-\theta)$ can be taken arbitrarily small, the right-hand member of this equation can be made as small as choose; that is, it can be taken less than an arbitrarily small positive member ϵ_1. We then have

$$\left| \int_C f(z)\, dz - \int_{C'} f(t)\, dt \right| < \epsilon_1 \int_C |dz| = \epsilon_1 \cdot L,$$

where L is the length of the curve C. But as L is finite, the product $\epsilon_1 \cdot L$ is arbitrarily small. The integral $\int_{C'} f(t)\, dt$ is zero by Theorem I. Hence we have

$$\int_C f(z)\, dz = 0.$$

THEOREM III. *Given a finite simply connected region S in which the integral $\int f(z)\, dz$ vanishes when taken along any closed curve C lying wholly within S. Any two paths of integration having the same extremities and lying wholly within S give the same value of the integral $\int f(z)\, dz$.*

Let $\alpha m\beta$, $\alpha n\beta$ be any two curves (Fig. 27) connecting the points α, β and lying wholly within the region S. We shall assume that these two curves do not intersect each other. The curve $\alpha m\beta$ followed by the curve $\beta n\alpha$ constitute a closed curve C. Then by hypothesis, we have

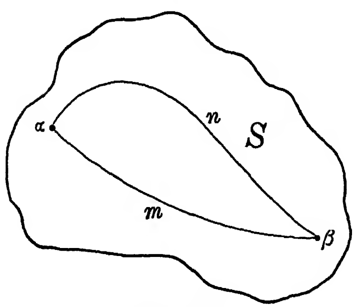

FIG. 27.

$$\int_C f(z)\, dz = \int_{\alpha m\beta} f(z)\, dz + \int_{\beta n\alpha} f(z)\, dz = 0,$$

or by reversing the direction in which the integral is taken along $\beta n\alpha$, we have by 1, Art. 17,

$$\int_{\alpha m\beta} f(z)\,dz - \int_{\alpha n\beta} f(z)\,dz = 0;$$

whence

$$\int_{\alpha m\beta} f(z)\,dz = \int_{\alpha n\beta} f(z)\,dz,$$

as required by the theorem.

It follows that, under the conditions set forth in the theorem, the value of the integral $\int f(z)\,dz$ depends upon the limits between which the integral is taken, but not upon the path. If one of these limits of integration, say β, is replaced by the variable z, the integral is a function of the upper limit and we may write

$$\int_{\alpha}^{z} f(z)\,dz = F(z).$$

Consequently, we have the following corollary.

Corollary I. *Given a finite simply connected region S in which $f(z)$ is holomorphic. The integral $\int f(z)\,dz$ taken along any two paths joining the same two fixed points of S and lying wholly within S has the same value for the two paths. If one limit of integration is a variable, the integral is then a function of that limit.*

This corollary is equivalent to saying that if $f(z)$ is holomorphic in a simply connected region S then a path of integration between two points of S can always be deformed into any other path lying wholly within S and joining the same two points without affecting the value of the integral.

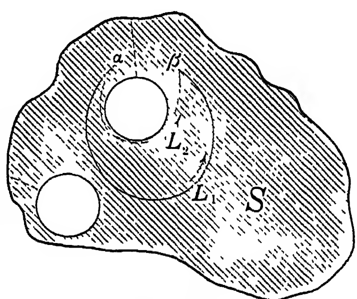

Fig. 28.

If the given region is not simply connected, it may be made so, as has been already pointed out, by the proper insertion of cross-cuts (Fig. 28). Of course the cross-cuts then form a part of the boundary and may not be crossed by the path of integration. In the resulting region the foregoing proposition applies, and consequently we have the following generalization of the foregoing corollary, namely:

Theorem IV. *Let S be any region in which $f(z)$ is holomorphic except at most at certain points. If a path of integration of $\int f(z)\,dz$*

between any two distinct points α, β of S is deformed into any other path between these same points such that it lies wholly in S, the value of the integral is not affected, provided that in the continuous deformation of the one path into the other no singular point of $f(z)$ is encountered.

If the path of integration is closed, that is if α and β become coincident, it follows that we may replace any such path C_1 (Fig. 29) by any other closed path C_2, provided that in the region S' bounded by the two curves no singularity of $f(z)$ is to be found; for, by introducing a cross-cut as shown in the figure, the curve C_2 together with the cross-cut may be regarded as obtained by means of a continuous deformation of C_1. But as the cross-cut is traversed twice, once in each direction, in taking the integral along the closed path from α back to the same point, it may be omitted. For a closed path of integration, we have therefore the following theorem.

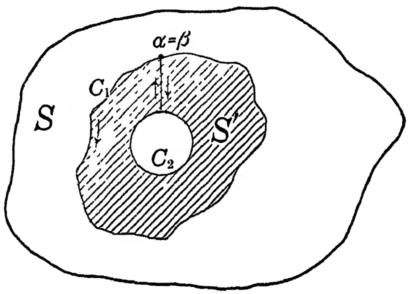

FIG. 29.

THEOREM V. *In any region S in which $f(z)$ is holomorphic except at certain points, a closed path C_1 of integration $\int f(z)\,dz$ may be replaced by any other closed path C_2 lying either interior or exterior to C_1, but lying wholly within S, provided the region bounded by the two curves incloses no singular points of $f(z)$ and no points not belonging to S.*

THEOREM VI. *If $f(z)$ is holomorphic in a finite closed multiply connected region S bounded by an exterior curve C and a finite number of inner curves $c_1, \ldots, c_n$,* then

$$\int_C f(z)\,dz = \sum_{k=1}^{n} \int_{c_k} f(z)\,dz,$$

each integral being taken in a positive direction with respect to the region inclosed.

Connect each inner curve c_k with the exterior curve C by a cross-cut, thus making the region simply connected. From Theorem I, the integral taken around the complete boundary, including the cross-cuts, is zero. However, it was shown in Art. 17 that this inte-

gral is equal to the integral taken over the boundary of the multiply connected region. We have then

$$\int_C f(z)\,dz + \sum_{k=1}^{n} \int_{c_k} f(z)\,dz = 0. \tag{15}$$

If now the integrals along the curves c_k are taken in a positive direction with respect to the regions interior to these curves rather than to the region S, the direction in which each integral is taken is changed, and hence by Theorem I, Art. 17, we have

$$\int_C f(z)\,dz = \sum_{k=1}^{n} \int_{c_k} f(z)\,dz,$$

as stated in the theorem.

20. Cauchy's integral formula. The following theorem, known as Cauchy's integral formula, is of fundamental importance, as it enables us to express the value of a function of a complex variable at any inner point of a finite closed region in which it is holomorphic, in terms of an integral taken around the boundary.

THEOREM I. *Given a finite closed region S whose boundary C consists of a finite number of ordinary curves. If $f(z)$ is holomorphic within S and converges uniformly along C, or if it is also holomorphic for values along C, then for any inner point α of S we have*

$$f(\alpha) = \frac{1}{2\pi i} \int_C \frac{f(z)\,dz}{z - \alpha}.$$

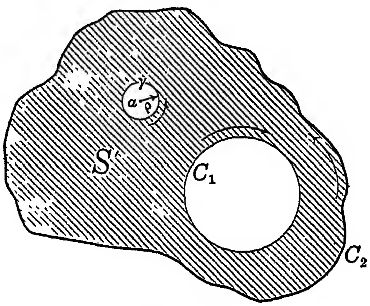

FIG. 30.

In accordance with the statement of the theorem, the boundary C may consist of one or more closed curves. For example, in Fig. 30 the complete boundary consists of the two curves C_1 and C_2. About the point α as a center, draw a circle γ of radius ρ lying entirely within the region S and inclosing only points of S. Denote by S' that portion of the region S lying outside the circle γ. Then the function $\dfrac{f(z)}{z - \alpha}$ is holomorphic in the region S', since by hypothesis

$f(z)$ is holomorphic in the region S. From Theorem I, Art. 19, we have upon integrating about the contour of S'

$$\int_C \frac{f(z)\,dz}{z-\alpha} + \int_\gamma \frac{f(z)\,dz}{z-\alpha} = 0,$$

or taking the integral about γ in a positive direction with respect to the region inclosed by γ, we have after transposing this integral to the second member of the equation

$$\int_C \frac{f(z)\,dz}{z-\alpha} = \int_\gamma \frac{f(z)\,dz}{z-\alpha}. \tag{1}$$

This result holds for all values of ρ, provided the circle lies entirely within S and incloses only points of S as stated above.

Because of the continuity of $f(z)$ at the point $z = \alpha$, we have

$$|f(z) - f(\alpha)| < \epsilon, \qquad |z - \alpha| \leqq \delta. \tag{2}$$

For an arbitrarily small ϵ, let z take values upon the circle γ whose radius ρ is not greater than δ. Consider now the integral

$$\int_\gamma \frac{f(z)\,dz}{z-\alpha} = f(\alpha)\int_\gamma \frac{dz}{z-\alpha} + \int_\gamma \frac{f(z)-f(\alpha)}{z-\alpha}\,dz. \tag{3}$$

From the Ex. of Art. 18, we have then

$$\int_\gamma \frac{dz}{z-\alpha} = 2\pi i. \tag{4}$$

To evaluate the second integral in the right-hand member of (3) put

$$z - \alpha = \rho(\cos\theta + i\sin\theta).$$

We then obtain

$$\begin{aligned}\int_\gamma \frac{f(z)-f(\alpha)}{z-\alpha}\,dz &= \int_\gamma \{f(z)-f(\alpha)\}\frac{dz}{z-\alpha}\\ &= \int_0^{2\pi}\{f(z)-f(\alpha)\}i\,d\theta. \end{aligned}\tag{5}$$

By 5, Art. 17, we have

$$\left|\int_\gamma \frac{f(z)-f(\alpha)}{z-\alpha}\,dz\right| \leqq \int_0^{2\pi} |f(z)-f(\alpha)|\cdot|d\theta|. \tag{6}$$

Since ρ, the radius of the circle of integration, was taken less than or at most equal to δ, we have by aid of (2)

$$\int_0^{2\pi}|f(z)-f(\alpha)|\cdot|d\theta| < \epsilon\int_0^{2\pi}|d\theta| = 2\pi\epsilon.$$

From (6), we have

$$\left|\int_\gamma \frac{f(z) - f(\alpha)}{z - \alpha}\, dz\right| < 2\pi\epsilon. \tag{7}$$

The foregoing integral has the same value if γ is any circle about α as a center, provided of course that the circle lies wholly within S. In other words, this integral is independent of the radius ρ and hence is independent of ϵ. Since $2\pi\epsilon$ is arbitrarily small, we have therefore

$$\int_\gamma \frac{f(z) - f(\alpha)}{z - \alpha}\, dz = 0. \tag{8}$$

Substituting the results given in (4) and (8) in equation (3), we have

$$\int_\gamma \frac{f(z)\, dz}{z - \alpha} = f(\alpha) \cdot 2\pi i.$$

Finally, we have from equation (1)

$$\int_C \frac{f(z)\, dz}{z - \alpha} = f(\alpha) \cdot 2\pi i,$$

or

$$f(\alpha) = \frac{1}{2\pi i}\int_C \frac{f(z)\, dz}{z - \alpha}, \tag{9}$$

which is the desired result.

As a consequence of the foregoing theorem, it is of importance to observe that the values of a function $f(z)$, which is holomorphic in a finite closed region S, are fully determined for values within S if we know its values upon the contour of that region.

The following theorem is of importance in the further development of the theory of analytic functions.

THEOREM II. *If $f(z)$ is holomorphic in a given finite region S, then the derivative $f'(z)$ is a continuous function in S; moreover, $f'(z)$ is itself holomorphic in S.*

Let z_0 be any inner point of S and let C be an ordinary closed curve, or combination of such curves, lying in S and forming a complete boundary having likewise the point z_0 as an inner point. For example, in Fig. 30 the complete boundary C consists of the two curves C_1 and C_2. From Theorem I, we have

$$f(z_0) = \frac{1}{2\pi i}\int_C \frac{f(t)\, dt}{t - z_0},$$

where t is a complex variable taken along the contour C.

Let $z_0 + \Delta z$ be any second point in the neighborhood of z_0, say within the circle γ lying within C and having z_0 as a center and ρ as a radius. We have then

$$\frac{f(z_0+\Delta z)-f(z_0)}{\Delta z} = \frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{(t-z_0-\Delta z)\,\Delta z} - \frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{(t-z_0)\,\Delta z}$$

$$= \frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{(t-z_0-\Delta z)\,(t-z_0)}. \tag{10}$$

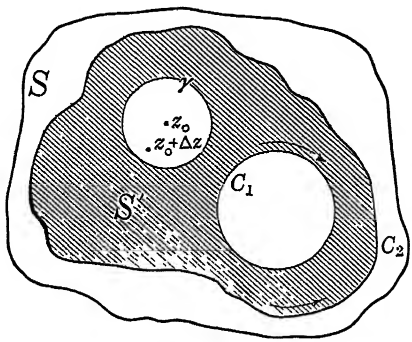

FIG. 31.

However,

$$\frac{1}{(t-z_0)\,(t-z_0-\Delta z)} = \frac{1}{(t-z_0)^2} + \frac{\Delta z}{(t-z_0)^2\,(t-z_0-\Delta z)},$$

and consequently,

$$\int_C \frac{f(t)\,dt}{(t-z_0)\,(t-z_0-\Delta z)} = \int_C \frac{f(t)\,dt}{(t-z_0)^2} + \int_C \frac{\Delta z f(t)\,dt.}{(t-z_0)^2\,(t-z_0-\Delta z)}. \tag{11}$$

We can readily show that the last of these integrals has the limit zero as $\Delta z \doteq 0$. To do so, let r be the lower limit of the distance of any point within γ from a point on C. We have then

$$|\,t-z_0-\Delta z\,| > r, \qquad |\,t-z_0\,| > r.$$

By use of **7**, Art. 17, we may now write

$$\left|\int_C \frac{\Delta z f(t)\,dt}{(t-z_0)^2\,(t-z_0-\Delta z)}\right| \leqq \frac{ML}{r^3}\,|\,\Delta z\,|,$$

where M is the maximum value of $|\,f(t)\,|$ along C and L is the length of the curve C. Hence as Δz approaches zero, we have zero as the limit of this integral.

Consequently, passing to the limit as $\Delta z \doteq 0$, we have from (11)

$$\underset{\Delta z \doteq 0}{L} \int_C \frac{f(t)\,dt}{(t-z_0)(t-z_0-\Delta z)} = \int_C \frac{f(t)\,dt}{(t-z_0)^2}.$$

Hence, from (10) we get

$$\underset{\Delta z \doteq 0}{L} \frac{f(z_0+\Delta z)-f(z_0)}{\Delta z} = \frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{(t-z_0)^2},$$

or

$$f'(z_0) = \frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{(t-z_0)^2}.$$

In the same way, we may show that

$$f''(z_0) = \frac{2!}{2\pi i}\int_C \frac{f(t)\,dt}{(t-z_0)^3},$$

$$f'''(z_0) = \frac{3!}{2\pi i}\int_C \frac{f(t)\,dt}{(t-z_0)^4},$$

$$\cdots\cdots\cdots\cdots$$

$$f^{(n)}(z_0) = \frac{n!}{2\pi i}\int_C \frac{f(t)\,dt}{(t-z_0)^{n+1}}.$$

The existence of these integrals enables us to affirm the existence of the higher derivatives of $f(z)$. Consequently, the derivative $f'(z)$ is continuous and holomorphic as the theorem states. A similar statement may now be made with reference to each of the higher derivatives.

Since $f'(z)$ is holomorphic in S if $f(z)$ is holomorphic in S, it follows that both $f(z)$ and $f'(z)$ are continuous in any closed region S' lying within S and hence by Theorem III, Art. 13, both $f(z)$ and $f'(z)$ are uniformly continuous in S'.

The fact that the continuity of the derivative $f'(z)$ follows from its existence renders the theory of analytic functions of a complex variable in many respects simpler than the theory of functions of a real variable; for, a derivative of a function of a real variable may exist at every point in an interval and yet not be continuous throughout the interval. In the next article, it will be shown that the continuity of the partial derivatives of the first order of u, v, where $f(z) = u(x, y) + iv(x, y)$, follow from the continuity of $f'(z)$. When that result has been established, we shall be able to apply to subsequent discussions the results of Green's theorem.

THEOREM III. *Let $f(t)$ be a continuous function of the complex variable t along an ordinary curve C, which may be closed or not. The integral*

$$\int_C \frac{f(t)\,dt}{t-z}$$

defines a function of z which is holomorphic for all values of z different from t.

It is at once evident that the given integral defines a function of z. We may put

$$F(z) = \int_C \frac{f(t)\,dt}{t-z}.$$

Then, by the reasoning employed in the demonstration of Theorem II, we obtain

$$F'(z) = \int_C \frac{f(t)\,dt}{(t-z)^2};$$

that is, $F(z)$ has a derivative for each value of z different from t. Hence, the function $F(z)$ is holomorphic for all such values of z and if $F(z)$ has any singular points, they must be points on the curve C.

By aid of the foregoing theorems, we can now prove the converse of the Cauchy-Goursat theorem. This theorem is due to Morera * and may be stated as follows:

THEOREM IV. *If $f(z)$ is continuous in a given region S and if the integral $\int f(z)\,dz$ is zero when taken around the complete boundary C of any portion of S, such that C lies wholly within S and incloses only points of S, then $f(z)$ is holomorphic in S.*

If the given region S is multiply connected, let it be made simply connected by the introduction of cross-cuts. Then every closed curve C lying within the new region S' is a complete boundary and by hypothesis the integral taken along such a curve is zero. We shall show that in this simply connected region $f(z)$ is holomorphic and hence holomorphic in S, even though this given region is multiply connected. Let α be a fixed point of S and z_0 any other point of the same region. Denote by $z_0 + \Delta z$ any point of S' in the neighborhood of z_0. Because

$$\int_C f(z)\,dz = 0,$$

* See *Reale Instituto Lombardo di scienze e lettere, Rendiconti*, 2 series, Vol. 19 (1886), p. 304.

it follows from Theorem III, Art. 19, that the value of the integral between any two points is independent of the path. Since we are at liberty, therefore, to select arbitrarily the path of integration between α and $z_0 + \Delta z$, without affecting the value of the integral $\int_\alpha^{z_0+\Delta z} f(z)\,dz$, we take a path passing through z_0 and rectilinear between z_0 and $z_0 + \Delta z$, as indicated in Fig. 32. The value of the integral $\int_\alpha^\zeta f(z)\,dz$, where ζ is a point upon this path, is a function of ζ, and we may write

$$F(\zeta) = \int_\alpha^\zeta f(z)\,dz.$$

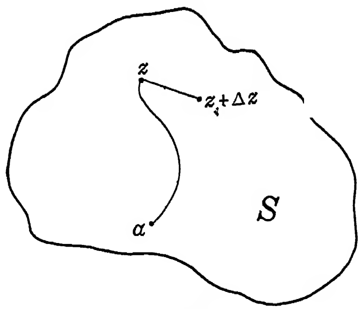

FIG. 32.

Hence, we have for $\zeta = z_0$ and $\zeta = z_0 + \Delta z$ the following values of $F(\zeta)$:

$$F(z_0) = \int_\alpha^{z_0} f(z)\,dz,$$

$$F(z_0 + \Delta z) = \int_\alpha^{z_0+\Delta z} f(z)\,dz,$$

whence

$$\Delta F(z_0) \equiv F(z_0 + \Delta z) - F(z_0) = \int_\alpha^{z_0+\Delta z} f(z)\,dz - \int_\alpha^{z_0} f(z)\,dz$$

$$= \int_{z_0}^{z_0+\Delta z} f(z)\,dz. \tag{12}$$

The given function $f(z)$ is continuous in S, and, therefore, we have for an arbitrarily small positive number ϵ, another positive number δ such that

$$|f(z) - f(z_0)| < \epsilon, \text{ for } |z - z_0| \equiv |\Delta z| < \delta.$$

This relation may be written

$$f(z) = f(z_0) + \eta(z),$$

where

$$|\eta| < \epsilon, \text{ for } |\Delta z| < \delta.$$

Putting $f(z_0) + \eta$ in place of $f(z)$ we obtain from (12)

$$\Delta F = \int_{z_0}^{z_0+\Delta z} f(z_0)\,dz + \int_{z_0}^{z_0+\Delta z} \eta\,dz. \tag{13}$$

Dividing by Δz, we have

$$\frac{\Delta F}{\Delta z} = \frac{1}{\Delta z}\int_{z_0}^{z_0+\Delta z} f(z_0)\, dz + \frac{1}{\Delta z}\int_{z_0}^{z_0+\Delta z} \eta\, dz. \tag{14}$$

But

$$\frac{1}{\Delta z}\int_{z_0}^{z_0+\Delta z} f(z_0)\, dz = \frac{f(z_0)}{\Delta z}\int_{z_0}^{z_0+\Delta z} dz,$$

which is equal to

$$\frac{f(z_0)}{\Delta z} \cdot \Delta z = f(z_0).$$

Moreover, we have

$$\left|\frac{1}{\Delta z}\int_{z_0}^{z_0+\Delta z} \eta\, dz\right| \leqq \frac{1}{|\Delta z|}\int_{z_0}^{z_0+\Delta z} |\eta| \cdot |dz|$$
$$< \frac{\epsilon}{|\Delta z|}\int_{z_0}^{z_0+\Delta z} |dz| = \epsilon.$$

From (14) we now get

$$\left|\frac{\Delta F}{\Delta z} - f(z_0)\right| < \epsilon, \qquad |\Delta z| < \delta;$$

that is,

$$\underset{\Delta z \doteq 0}{L} \frac{\Delta F}{\Delta z} \equiv F'(z_0) = f(z_0).$$

Since z_0 was taken to be any point of S, it follows that for all values of z in S we have

$$F'(z) = f(z). \tag{15}$$

Consequently, $F(z)$ is holomorphic in S; but, as we have seen, the derivative of such a function is also holomorphic; hence $f(z)$ must likewise be holomorphic in S, as the theorem requires.

The Cauchy-Goursat theorem states a necessary condition that a given function $f(z)$ shall be holomorphic in a given region. The theorem just demonstrated gives a sufficient condition that a continuous function is holomorphic. We may combine these two results into the following theorem.

THEOREM V. *The necessary and sufficient condition that a continuous function $f(z)$ is holomorphic in a given finite region S is that the integral $\int f(z)\, dz$ is zero when taken along the complete boundary C of any portion of the plane when C lies entirely within S and incloses only points of S.*

21. Cauchy-Riemann differential equations. In Art. 20 we discussed the necessary and sufficient condition that a function $f(z)$

is holomorphic in a given finite region S. This condition was expressed in terms of a definite integral. It is often more convenient to have such a condition expressed in terms of the partial derivatives of u and v, where

$$f(z) \equiv w = u(x, y) + iv(x, y)$$

is the given function. Such a criterion is given in the following theorem.

THEOREM I. *In a given finite region S let u and v be two real single-valued functions of the real variables x, y. The necessary and sufficient condition that the complex function*

$$w = u + iv$$

is holomorphic in S is that the partial derivatives of u and v of the first order exist and are continuous and moreover satisfy the following partial differential equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{1}$$

These differential equations are known as the **Cauchy-Riemann differential equations.** To show that these equations present a necessary condition we proceed as follows. Since the function w is holomorphic in S, it has a derivative with respect to z. As we have seen, the existence of this derivative involves the condition that the ratio $\frac{\Delta w}{\Delta z}$ shall have the same limiting value as Δz approaches zero in any direction whatsoever. Consequently, the same limiting value is obtained if Δz is permitted to approach zero through real values or through purely imaginary values. The increment $\Delta z = \Delta x + i\,\Delta y$ becomes in the first case Δx and in the second case $i\,\Delta y$. We may therefore write

$$\underset{\Delta z \doteq 0}{L} \frac{\Delta w}{\Delta z} = \underset{\Delta x \doteq 0}{L} \frac{\Delta w}{\Delta x} = \underset{\Delta y \doteq 0}{L} \frac{1}{i} \frac{\Delta w}{\Delta y}. \tag{2}$$

Since the first limit exists by hypothesis, the second and third limits must also exist. We have then

$$\frac{dw}{dz} = \frac{\partial w}{\partial x} = \frac{1}{i} \frac{\partial w}{\partial y}. \tag{3}$$

However, we have

$$\frac{\partial w}{\partial x} = \frac{\partial}{\partial x}[u(x, y) + iv(x, y)] = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}, \tag{4}$$

$$\frac{\partial w}{\partial y} = \frac{\partial}{\partial y}[u(x, y) + iv(x, y)] = \frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y}. \tag{5}$$

Substituting these values in (3), we obtain

$$\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}. \tag{6}$$

Equating the real parts and the imaginary parts in this equation, we have

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{7}$$

By Theorem II, Art. 20, the derivative $\frac{dw}{dz}$ is continuous. The continuity of the partial derivatives in (7) follows therefore from equations (3), (4), and (5) and Theorem II, Art. 13.

We may now show the conditions of the theorem to be sufficient as follows. From equation (2), Art. 17, we have

$$\int_C f(z)\,dz = \int_C u\,dx - v\,dy + i\int_C v\,dx + u\,dy, \tag{8}$$

where C may be regarded as any path of integration within the given region S. As u, v are continuous, both of the integrals

$$\int_C u\,dx - v\,dy, \qquad \int_C v\,dx + u\,dy \tag{9}$$

exist. By hypothesis, the equations (1) are satisfied by u, v. Hence, by Theorem II, Art. 16, both of the integrals in (9) are zero, and we have from (8)

$$\int_C f(z)\,dz = 0$$

for every path of integration C forming a complete boundary of any portion of S such that C lies entirely within S and incloses only points of S. Consequently, by Theorem IV, Art. 20, $f(z)$ is holomorphic in the given region S as the theorem requires.

From equations (3), (4), and (5), we have

$$\frac{dw}{dz} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}. \tag{10}$$

This relation affords a convenient method of computing the derivative of w, when $w = f(z)$ is expressed in terms of x and y.

Ex. Given

$$w = x^3 + 3x^2yi - 3xy^2 - y^3i.$$

Show that w is holomorphic everywhere in the finite region, and compute D_zw.

We have

$$u = x^3 - 3xy^2, \qquad v = 3x^2y - y^3,$$

$$\frac{\partial u}{\partial x} = 3x^2 - 3y^2, \qquad \frac{\partial v}{\partial x} = 6xy,$$

$$\frac{\partial u}{\partial y} = -6xy, \qquad \frac{\partial v}{\partial y} = 3x^2 - 3y^2.$$

The conditions of the theorem are fulfilled and the function is therefore holomorphic for all finite values of x and y. For the derivative D_zw we have

$$D_zw = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = 3x^2 - 3y^2 + 6ixy.$$

In this particular case w can be readily expressed directly as a function of z, namely $w = z^3$. We have then by the laws of differentiation $D_zw = 3z^2$, a result which is identical with that already obtained.

In the calculus of real variables and in trigonometry, we became acquainted with certain pairs of functions that we called **inverse functions**. For example, $y = \sin x$ and $x = \text{arc} \sin y$, $y = x^2$ and $x = \sqrt{y}$ are illustrations of inverse functions. If we have given any function $w = f(z)$, z may be regarded as a function of w, expressed implicitly by this equation in w and z. It is desirable, however, to know the conditions under which z, considered as a function of w, is holomorphic in a definite region when $f(z)$ is holomorphic in a given region. The desired conditions may be stated as follows:

THEOREM II. *Let $w = f(z)$ be holomorphic in a given region T, which is defined by the inequality $|z - z_0| < h$. Moreover, let $f'(z) \neq 0$ for values of z in T and let $w_0 = f(z_0)$. Then corresponding to the number h there is determined a number k such that in the W-plane there exists a region S, defined by the inequality $|w - w_0| < k$, for values of w within which the equation*

$$w = f(z),$$

has one and only one solution

$$z = F(w);$$

moreover the inverse function thus determined is holomorphic in S and

$$F'(w) = \frac{1}{f'(z)}.$$

We have

$$w = f(z) = u + iv.$$

The functions u, v have continuous first derivatives which satisfy simultaneously the equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{11}$$

Consider now the Jacobian

$$\begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}. \tag{12}$$

By aid of the conditional equations (11), this determinant can be written in the form

$$\begin{vmatrix} \dfrac{\partial u}{\partial x} & -\dfrac{\partial v}{\partial x} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial u}{\partial x} \end{vmatrix} = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 = \left|\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right|^2 = |\, f'(z)\,|^2.$$

By hypothesis $f'(z)$ is different from zero, and, hence, the Jacobian does not vanish. The non-vanishing of this determinant is, however, the condition that a region S of the W-plane exists, for every point of which the two equations

$$u = \Psi_1(x, y), \quad v = \Psi_2(x, y)$$

have one and only one simultaneous solution, expressing x and y in terms of u, v, say

$$x = \chi_1(u, v), \quad y = \chi_2(u, v),$$

where these new functions of u, v have continuous first derivatives with respect to u and to v.* We may then write

$$z = x + iy = \chi_1(u, v) + i\chi_2(u, v) = F(w).$$

It still remains, however, to show that the function

$$z = F(w),$$

found in this way, is holomorphic in the region S.

* See *Encyklopädie der Math. Wissenschaften*, Vol. II, Heft 1, Art. 5; also Goursat-Hedrick, *Mathematical Analysis*, Vol. I, p. 50.

We have for Δw, Δz, each different from zero,

$$\frac{\Delta z}{\Delta w} = \frac{1}{\dfrac{\Delta w}{\Delta z}}. \tag{13}$$

We wish to consider the limit as Δw approaches zero. Since $F(w)$ is continuous, we know that Δz approaches zero simultaneously with Δw. We may therefore write

$$\underset{\Delta w \doteq 0}{L} \frac{\Delta z}{\Delta w} = \underset{\Delta z \doteq 0}{L} \frac{1}{\dfrac{\Delta w}{\Delta z}}. \tag{14}$$

But as $f'(z) \equiv \underset{\Delta z \doteq 0}{L} \dfrac{\Delta w}{\Delta z}$ is different from zero, we have

$$\underset{\Delta w \doteq 0}{L} \frac{\Delta z}{\Delta w} = \underset{\Delta z \doteq 0}{L} \frac{1}{\dfrac{\Delta w}{\Delta z}} = \frac{1}{\underset{\Delta z = 0}{L} \dfrac{\Delta w}{\Delta z}}. \tag{15}$$

In this equation we know that the limit in the right-hand member exists and defines $\dfrac{1}{f'(z)}$, because $w = f(z)$ is by hypothesis holomorphic in the given region T. It follows from (13) that the limit $\underset{\Delta w \doteq 0}{L} \dfrac{\Delta z}{\Delta w}$ must also exist for all points in S; that is, $z = F(w)$ has a derivative for all values of w in S. Hence, $F(w)$ is holomorphic in S as the theorem requires.

From (15) we have the relation

$$F'(w) = \frac{1}{f'(z)}, \tag{16}$$

between the derivative of the given function and that of its inverse, which was also to be demonstrated.

In Art. 20, it was shown that when $f(z)$ is holomorphic, $f'(z)$ is necessarily continuous. It can now be demonstrated that the difference quotient converges uniformly to the derivative as stated in the following theorem.

THEOREM III. *If $f(z)$ is holomorphic in a finite closed region S, then the difference quotient*

$$\frac{f(z + \Delta z) - f(z)}{\Delta z}$$

converges uniformly to the limit $f'(z)$ for values of z in S.

This theorem is equivalent to saying that if $f(z)$ is holomorphic in

S then for an arbitrarily chosen positive number ϵ, there exists another positive number η such that

$$\left|\frac{f(z+\Delta z)-f(z)}{\Delta z}-f'(z)\right|<\epsilon, \qquad |\Delta z|<\eta \tag{17}$$

for all values of z in S. In other words, the value of ϵ being determined arbitrarily there exists a number η independent of z satisfying the required condition. We may put

$$z = x + iy, \qquad f(z) = u(x, y) + iv(x, y),$$

and hence have

$$f(z+\Delta z) - f(z) \equiv \Delta f(z) = \Delta u + i\,\Delta v, \qquad \Delta z = \Delta x + i\,\Delta y. \tag{18}$$

Since u is a function of x and y, we may write

$$\begin{aligned}\Delta u &= u(x+\Delta x, \quad y+\Delta y) - u(x, y)\\ &= u(x+\Delta x, \quad y+\Delta y) - u(x+\Delta x, y) + u(x+\Delta x, y) - u(x, y).\end{aligned}$$

By making use of the law of the mean, this result may be written

$$\begin{aligned}\Delta u = u_y'(x+\Delta x, y+\theta_1\,\Delta y)\,\Delta y + u_x'(x+\theta_2\,\Delta x, y)\,\Delta x,\\ 0<\theta_1<1, \qquad 0<\theta_2<1,\end{aligned} \tag{19}$$

where u_y', u_x' denote partial derivatives with respect to x and y, respectively. In a similar manner, we may get

$$\begin{aligned}\Delta v = v_y'(x+\Delta x, y+\theta_3\,\Delta y)\,\Delta y + v_x'(x+\theta_4\,\Delta x, y)\,\Delta x,\\ 0<\theta_3<1, \qquad 0<\theta_4<1.\end{aligned} \tag{20}$$

We have seen, under Theorem I, that

$$f'(z) = u_x'(x, y) + iv_x'(x, y).$$

By aid of (19) and (20) and the Cauchy-Riemann differential equations, we may now write

$$\begin{aligned}\frac{f(z+\Delta z)-f(z)}{\Delta z} - f'(z) &= \frac{f(z+\Delta z)-f(z)-\Delta z f'(z)}{\Delta z}\\ &= \frac{\Delta u + i\,\Delta v - (\Delta x + i\,\Delta y)(u_x' + iv_x')}{\Delta z}\\ &= [u_y'(x+\Delta x, y+\theta_1\,\Delta y) - u_y'(x,y)]\frac{\Delta y}{\Delta x + i\,\Delta y}\\ &\quad + [u_x'(x+\theta_2\,\Delta x, y) - u_x'(x, y)]\frac{\Delta x}{\Delta x + i\,\Delta y}\\ &\quad + i[v_y'(x+\Delta x, y+\theta_3\Delta y) - v_y'(x, y)]\frac{\Delta y}{\Delta x + i\Delta y}\\ &\quad + i[v_x'(x+\theta_4\,\Delta x, y) - v_x'(x, y)]\frac{\Delta x}{\Delta x + i\,\Delta y}.\end{aligned}$$

Since u_x', u_y', v_x', v_y' are continuous in S they are uniformly continuous in this region * and each of the expressions inclosed in brackets can consequently be made less in absolute value than $\frac{\epsilon}{4}$ by the proper choice of Δx, Δy; for example, if Δx, Δy be so taken that $\sqrt{\Delta x^2 + \Delta y^2} < \eta$. The absolute value of each of the factors outside of the brackets is never greater than unity. Consequently, we may write

$$\left|\frac{f(z + \Delta z) - f(z)}{\Delta z} - f'(z)\right| < \epsilon, \qquad \sqrt{\Delta x^2 + \Delta y^2} \equiv |\Delta z| < \eta.$$

Hence the theorem.

22. Change of complex variable. In Art. 18, it was pointed out that the law for the change of variable in a definite integral of a function of a real variable can be extended without modification to the case of a definite integral of a function of a complex variable where the change is from a complex variable to a real variable. We can now complete the discussion by showing that the same law holds where the change is from one complex variable to another.

Suppose we put

$$z = \phi(t),$$

where t is complex and $\phi(t)$ is holomorphic along a curve K. As t traces the curve K, suppose the variable z traces the path C. Corresponding to the points of division $t_0, t_1, \ldots, t_n$ upon K, we have the points $z_0, z_1, \ldots, z_n$ of division of C. Since $\phi(t)$ is holomorphic along K and hence has a derivative $\phi'(t)$ along K, we have

$$\frac{\Delta_k z}{\Delta_k t} = \frac{z_k - z_{k-1}}{t_k - t_{k-1}} = \phi'(t_{k-1}) + \epsilon_k, \qquad k = 0, 1, 2, \ldots n, \tag{1}$$

where ϵ_k vanishes with $\Delta_k t$. If in the sum

$$\sum f(z_{k-1}) \Delta_k z,$$

we replace $\Delta_k z$ by its value obtained from (1), we have

$$\sum f(z_{k-1}) \Delta_k z = \sum f(z_{k-1}) \phi'(t_{k-1}) \Delta_k t + \sum \epsilon_k f(z_{k-1}) \Delta_k t. \tag{2}$$

Since $z = \phi(t)$ is by hypothesis holomorphic along K, it follows from Theorem III, Art. 21, that corresponding to an arbitrarily small

* See Goursat-Hedrick, *Mathematical Analysis*, Vol. I, p. 251.

positive number ϵ there exists a positive number δ, such that for $|\Delta_k t| < \delta$ we have

$$\left|\frac{\Delta_k z}{\Delta_k t} - \phi'(t_{k-1})\right| \equiv |\epsilon_k| < \epsilon;$$

that is, the various values of ϵ_k can be replaced by a single arbitrarily small value ϵ, if the values of $\Delta_k t$ are all taken less in absolute value than δ. We then have

$$\left|\sum \epsilon_k f(z_{k-1})\,\Delta_k t\right| \leqq \epsilon M \sum |\Delta_k t| \leqq \epsilon ML, \tag{3}$$

where M is the maximum value of $|f(z)|$ upon C and L is the length of K. Since ML is a constant and ϵ is arbitrarily small the limit of the sum in (3) is zero as Δt approaches zero; hence, we have from (2) and (3) upon passing to the limit

$$\int_C f(z)\,dz = \int_K f\{\phi(t)\}\phi'(t)\,dt, \tag{4}$$

which expresses the law for the change of variable for the case under consideration.

Ex. Given a region S consisting of that portion of the complex plane exterior to its boundary C, which is taken to be a closed curve exterior to the unit circle about the origin. Consider the integral of $f(z) = \frac{1}{z^3}$ taken over the boundary C of the region S.

By putting $z = \frac{1}{z'}$, C goes into a curve K about the origin and lying within the unit circle. We have

$$\int_C f(z)\,dz = \int_C \frac{1}{z^3}\,dz = \int_K z'^3 \cdot \frac{-1}{z'^2}\,dz'$$

$$= -\int_K z'\,dz' = 0. \qquad \text{(Th. IV, Art. 20.)}$$

It will be observed that while the first integral must be taken in a clockwise direction, as the region S with respect to which the integral is taken lies exterior to the curve C, the integral along the curve K is taken in a counter-clockwise direction; for, as z traces out C in a clockwise direction $\frac{1}{z}$ traces out K in a counter-clockwise direction.

23. Indefinite integrals. Let $f(z)$ be holomorphic in a given finite region S. As we have seen, the integral $\int_\alpha^z f(z)\,dz$ defines in S a function $F(z)$ of the variable limit of integration. From the dis-

cussion of Theorem III of Art. 20, it follows that this function $F(z)$ is also holomorphic in S, and furthermore that the relation between $f(z)$ and $F(z)$ is such that

$$f(z) = \frac{dF(z)}{dz}.$$

Let $\phi(z)$ denote any function having the derivative $f(z)$. Such a function is called a **primitive function** of $f(z)$. We write as in the calculus of real variables

$$\phi(z) = \int f(z)\,dz,$$

and speak of $\int f(z)\,dz$ as the **indefinite integral.**

The primitive function $\phi(z)$ can differ from $F(z)$ at most by a constant. For, $\phi(z)$ is holomorphic in S, since it has a derivative $f(z)$, and hence $F(z) - \phi(z)$ is also holomorphic in S. We have, since both $\phi(z)$ and $F(z)$ are primitive functions of $f(z)$,

$$\frac{d}{dz}[F(z) - \phi(z)] = \frac{dF}{dz} - \frac{d\phi}{dz} = f(z) - f(z) = 0 \tag{5}$$

for all values of z in S. We may write

$$F(z) - \phi(z) = u(x, y) + iv(x, y),$$

and have from Art. 21

$$\frac{d}{dz}[F(z) - \phi(z)] = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}. \tag{6}$$

From (5) and (6) it follows that

$$\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = 0,$$

and hence we must have

$$\frac{\partial u}{\partial x} = 0, \qquad \frac{\partial v}{\partial x} = 0. \tag{7}$$

From the Cauchy-Riemann differential equations, we have also

$$\frac{\partial u}{\partial y} = 0, \qquad \frac{\partial v}{\partial y} = 0. \tag{8}$$

As (7) and (8) hold for all values of (x, y) in S, it follows that both u and v are real constants.* Consequently, $u + iv$ must be a complex

* Compare Pierpont, *Theory of Functions of Real Variables*, Vol. I, p. 250.

constant and $F(z)$ differs from $\phi(z)$ by this constant value for all values of z in S.

Since $\phi(z)$ differs from $F(z)$ by a constant for all values of z along the path of integration between any two points α, β of S, we may write

$$\phi(z) = \int_\alpha^z f(z)\,dz + c.$$

For $z = \alpha$, this relation becomes

$$\phi(\alpha) = c.$$

If, on the other hand, $z = \beta$, we get

$$\phi(\beta) = \int_\alpha^\beta f(z)\,dz + c.$$

From these two results, we have at once the fundamental theorem of the integral calculus, namely:

$$\int_\alpha^\beta f(z)\,dz = \phi(\beta) - \phi(\alpha); \tag{9}$$

that is, the law for the evaluation of a definite integral in the calculus of real variables may be extended without modification to functions which are holomorphic in a given finite region.

Ex. Given the function $f(z) = z^m$, $m \neq -1$; find the value of the integral $\int_{z_0}^{z_1} f(z)\,dz$.

We have $$\phi(z) = \int f(z)\,dz = \frac{z^{m+1}}{m+1} + c.$$

Hence, from (1) we obtain

$$\int_{z_0}^{z_1} f(z)\,dz = \left.\frac{z^{m+1}}{m+1}\right]_{z_0}^{z_1} = \frac{z_1^{m+1} - z_0^{m+1}}{m+1}.$$

24. Laplace's differential equation. We have the following theorem.

THEOREM I. *In a given finite region S, let the complex function*

$$f(z) = u + iv$$

be holomorphic; then the functions $u(x, y)$, $v(x, y)$ satisfy the partial differential equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{1}$$

This differential equation is known as **Laplace's differential equation** and is of prime importance in theoretical physics.

Since $f(z)$ is holomorphic in S, the derivative $f'(z)$ and also the higher derivatives exist and are holomorphic in the same region. It follows that the partial derivatives of u, v with respect to x and y exist and are continuous. This statement holds not only for the partial derivatives of the first order but likewise for those of the second and higher orders. From the Cauchy-Riemann differential equations,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}, \tag{2}$$

we obtain by differentiation

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x\, \partial y}, \qquad \frac{\partial^2 u}{\partial y^2} = -\frac{\partial^2 v}{\partial y\, \partial x}. \tag{3}$$

As the partial derivatives of the second order are continuous in x, y, together, we have *

$$\frac{\partial^2 v}{\partial x\, \partial y} = \frac{\partial^2 v}{\partial y\, \partial x}. \tag{4}$$

Hence, by addition of the equations (3), we have

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{5}$$

In a similar manner we can show that

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0.$$

We shall now consider the converse proposition, namely:

THEOREM II. *If in a given finite region S, a function $u(x, y)$ has continuous partial derivatives of the first and second order and satisfies Laplace's differential equation, then there exists a function $v(x, y)$ determined except as to an additive constant, such that the complex function $u + iv = f(z)$ is holomorphic in S.*

We have given the condition that u satisfies the differential equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

* See Townsend and Goodenough, *First Course in Calculus*, Art. 104.

We now define v by the relation

$$v = \int_{x_0, y_0}^{x, y} - \frac{\partial u}{\partial y} dx + \frac{\partial u}{\partial x} dy + C. \tag{6}$$

This integral exists because the integrand is continuous, since $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$ are continuous. Moreover, the integral is independent of the path by virtue of Theorem III, Art. 16, because

$$\frac{\partial}{\partial y}\left(-\frac{\partial u}{\partial y}\right) = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right);$$

that is,

$$\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial x^2} = 0.$$

Differentiating v partially with respect to x, we have from (6) by aid of Theorem IV, Art. 16,

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \tag{7}$$

Similarly, differentiating with respect to y, we get

$$\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x}. \tag{8}$$

Equations (7), (8) are however none other than the Cauchy-Riemann differential equations, and hence

$$u + iv = f(z)$$

is holomorphic in S.

Ex. 1. Given $u = x^3 - 3xy^2$. Show that there exists a function $v(x, y)$ such that $w = u + iv$ is holomorphic in the finite region. Determine the function $v(x, y)$.

The given function satisfies Laplace's equation; for, we have

$$\frac{\partial u}{\partial x} = 3x^2 - 3y^2, \qquad \frac{\partial u}{\partial y} = -6xy,$$
$$\frac{\partial^2 u}{\partial x^2} = 6x, \qquad \frac{\partial^2 u}{\partial y^2} = -6x,$$

hence

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

The required value of v is to be determined from equation (6). As pointed out in the discussion of the foregoing theorem, this integral is independent of the path. It can be conveniently evaluated by making the path rectilinear passing

from (x_0, y_0) to (x, y_0), thence to (x, y). From (x_0, y_0) to (x, y_0) we have $y = y_0$, $dy = 0$, while x varies from x_0 to x. From (x, y_0) to (x, y), we have $dx = 0$ and y varying from y_0 to y. Hence, from equation (6) we have

$$\begin{aligned} v &= \int_{x_0}^{x} 6\,xy_0\,dx + \int_{y_0}^{y} (3\,x^2 - 3\,y^2)\,dy + C \\ &= 3\,x^2y_0 - 3\,x_0^2y_0 + 3\,x^2y - y^3 - 3\,x^2y_0 + y_0^3 + C \\ &= 3\,x^2y - y^3 - (3\,x_0^2y_0 - y_0^3) + C. \end{aligned}$$

Putting
$$C - 3\,x_0^2y_0 + y_0^3 = c,$$
we have
$$v = 3\,x^2y - y^3 + c,$$
whence
$$\begin{aligned} f(z) = u + iv &= x^3 - 3\,xy^2 + i\,\{(3\,x^2y - y^3) + c\} \\ &= (x + iy)^3 + ic = z^3 + ic. \end{aligned}$$

From the form of $f(z)$, it will be at once seen that it is holomorphic in any finite region.

Ex. 2. Given $u = \log\,(x^2 + y^2)^{\frac{1}{2}}$. Find a function $v(x, y)$ such that $u + iv$ is an analytic function.

We have

$$\frac{\partial u}{\partial x} = \frac{x}{(x^2 + y^2)}, \qquad \frac{\partial u}{\partial y} = \frac{y}{(x^2 + y^2)},$$
$$\frac{\partial^2 u}{\partial x^2} = \frac{y^2 - x^2}{(x^2 + y^2)^2}, \qquad \frac{\partial^2 u}{\partial y^2} = \frac{x^2 - y^2}{(x^2 + y^2)^2}.$$

These results substituted in Laplace's equation show that u satisfies that equation. As in Ex. 1, it is convenient to take the path as rectilinear through the intermediate point (x, y_0). From equation (6), we then have

$$\begin{aligned} v &= \int_{x_0}^{x} \frac{-y_0}{x^2 + y_0^2}\,dx + \int_{y_0}^{y} \frac{x}{x^2 + y^2}\,dy + C \\ &= -\text{arc tan}\,\frac{x}{y_0} + \text{arc tan}\,\frac{x_0}{y_0} + \text{arc tan}\,\frac{y}{x} - \text{arc tan}\,\frac{y_0}{x} + C \\ &= \text{arc tan}\,\frac{y}{x} - \frac{\pi}{2} + \text{arc tan}\,\frac{x_0}{y_0} + C. \end{aligned}$$

If we now put
$$C - \frac{\pi}{2} + \text{arc tan}\,\frac{x_0}{y_0} = c,$$
we have
$$v = \text{arc tan}\,\frac{y}{x} + c,$$
and hence get
$$u + iv = \log\,(x^2 + y^2)^{\frac{1}{2}} + i\,\text{arc tan}\,\frac{y}{x} + ic.$$

The function
$$w = f(z) = u + iv$$
is holomorphic for all values of z in any finite region not including the origin, since for such values of $z \equiv x + iy$ the functions u and v satisfy the conditions of Theorem II.

25. Applications to physics. A variety of problems in mathematical physics are associated with the solution of Laplace's equation. According to Newton's law two bodies in space attract each other directly as the product of their masses and inversely as the square of the distance between them. When one of these bodies is moved with respect to the other, then work is done in overcoming the attractive force of the second body. The work done in overcoming the attractive force of a given mass M so as to move a particle of unit mass from a given point to an infinite distance is defined as the **potential** of M at that point. It can be shown that the potential is a function of the space coördinates x, y, z alone; that is, that it is independent of the path. A Newtonian potential function $u(x, y, z)$ due to attractive matter is such that Laplace's equation,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0,$$

must be satisfied whenever (x, y, z) are the coördinates of a point exterior to the matter itself.

If the conditions are such that the attractive force acts only in a plane, taken conveniently as the XY-plane, then the third component of the force becomes zero and $\frac{\partial^2 u}{\partial z^2}$ vanishes. Consequently, for two dimensions Laplace's equation takes the form discussed in the preceding article, namely

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

In this case the potential is a **logarithmic potential,** and the force overcome varies directly as the product of the masses of the particles and inversely as the distance between them.

From the discussion in the last article, it follows that if u is a logarithmic potential, there exists a function v, determined except as to an additive constant, such that

$$w = u(x, y) + iv(x, y),$$

considered as a function of z, is holomorphic in the region for which the potential is determined.

A potential also exists in connection with a magnetic or electric field. An **electric potential** at any point may be defined as the work

necessary to be done against an electric force in moving a unit charge of negative electricity from that point to an infinite distance. The potential may be defined in a similar manner for the points of a magnetic field. In any case the derivatives of the potential with respect to x, y represent the components of the force in the direction of the two axes. In order to have the two-dimensional case that arises in connection with the discussion of functions of a complex variable, the force exerted must be confined to a plane, taken as the complex plane. For example, such a case arises when a current of electricity flows through a straight wire of indefinite length. A magnetic field is created in the surrounding space such that the component of the force in the direction of the wire is zero. Consequently any plane perpendicular to the wire may be taken as the complex plane and the application reduces to one in two dimensions.

As another illustration of the applications of the functions of a complex variable may be mentioned the stationary streaming of electricity. Suppose, for example, we have given as a conductor a thin sheet of metal of unlimited extent and of uniform thickness and structure. Let the current of electricity be introduced into and leave this conductor by means of perfectly conducting electrodes. The current may then be regarded as flowing in a plane parallel to the two surfaces of the sheet. The illustration then becomes a two-dimensional one and the condition that the streaming is stationary is that for each value of (x, y) in the region of flow the potential is such that Laplace's equation for two dimensions is satisfied.*

A corresponding application to the flow of heat may be readily formulated. Let the body in which the flow takes place be a cylinder of indefinite length whose rectangular cross-section consists of one or more closed curves. Upon the surface of this cylinder let the temperature u be a constant for all points along the same generator of the cylinder. Moreover, let the temperature along any line parallel to a generator and lying within the cylinder be constant. Otherwise let the temperature vary continuously both upon the surface and within the cylinder. The flow of heat then takes place in planes perpendicular to the generators of the cylinder. The temperature u must satisfy Laplace's equation for two dimensions, if the flow is continuous.

The last two illustrations are special cases of the flow of incompres-

* See Jeans, *Electricity and Magnetism*, Art. 389.

sible fluids. If u is a function of the space coördinates x, y, z such that the components of the velocity of the fluid are

$$-\frac{\partial u}{\partial x}, \qquad -\frac{\partial u}{\partial y}, \qquad -\frac{\partial u}{\partial z},$$

then u is called a **velocity-potential** * in analogy to the Newtonian potential function already discussed. The existence of a velocity-potential is a property not of a region of space but of portions of matter. As the portion of matter moves about, it carries this property with it, while the space occupied by the matter at any instant may come to be occupied by matter not possessing the property. An irrotational motion of a fluid within a simply connected region is characterized by the existence of a velocity-potential. The condition that the given fluid flows continuously and has a velocity-potential u is that u satisfies Laplace's differential equation. If we now impose such conditions upon the fluid that the flow takes place in a plane, then $\frac{\partial u}{\partial z}$ vanishes and the problem reduces to a two-dimensional one, and the theory of functions of a complex variable may be applied. To accomplish this purpose, let us suppose that the fluid is of constant density and flows between two fixed parallel planes so that the path of the individual points of the fluid lies in a plane parallel to the fixed planes and the fluid flows so that two points which at any instant lie in a line perpendicular to the fixed planes remain in the same relative position. Then any plane parallel to the fixed planes may be taken as the complex plane and the theory of functions of a complex variable becomes at once applicable.

In the next chapter we shall discuss more in detail some examples illustrating the way functions of a complex variable may be employed in particular physical problems.

EXERCISES

1. Show that

$$w = x^4 + 4ix^3y - 6x^2y^2 - 4ixy^3 + y^4$$

satisfies the conditions given in Art. 21 for finite values of x and y and is therefore holomorphic in the finite region. Express w in terms of z and compute $D_z w$ by the method given in that article.

2. Given $f(z) = z^n$, where n is a positive integer. Find $f(z_0 + \Delta z) - f(z_0)$ by means of the binomial theorem and show for all values of z, (a) that $f(z)$ is continuous, (b) that $f'(z) = nz^{n-1}$, making use of the definition of a derivative.

* For fuller discussion of the properties of velocity-potentials, see Lamb, *Hydrodynamics*, 3d Ed., Chapters II and III.

3. By making use of the definition of a derivative and the methods employed in the calculus of real variables prove that, if f, f_1, f_2 are holomorphic in a region R and hence each has a derivative in this region, the following laws hold for values of z in R:

(*a*) If f is a constant, then $f' = 0$.
(*b*) If $f = f_1 \pm f_2$, then $f' = f_1' \pm f_2'$.
(*c*) If $f = f_1 \cdot f_2$, then $f' = f_1 \cdot f_2' + f_2 \cdot f_1'$.
(*d*) If $f = \dfrac{f_1}{f_2}$, where $f_2 \neq 0$ for all values of the argument considered, then

$$f' = \frac{f_2 \cdot f_1' - f_1 \cdot f_2'}{(f_2)^2}.$$

4. Show that every rational integral function of z is holomorphic in the entire finite portion of the complex plane, also that every rational function of z is holomorphic in any region not including points where the denominator is zero.

5. Find the value of the line-integral

$$\int_C (3\,x + 7\,y^2)\,dx + (x^2 + 3\,y)\,dy,$$

where C is the perimeter of a square whose sides are $x = 0, x = 4, y = 2, y = -2$. Is this line-integral independent of the path?

6. Evaluate the line-integral

$$\int_C (x^2 + 7\,xy)\,dx + (3\,x + y^2)\,dy,$$

where C is the boundary of the multiply connected region bounded by the two curves whose equations are $x^2 + y^2 = 9$, $(x + 1)^2 + y^2 = 1$.

7. Let the path C of integration be given by the equations

$$x = 3\cos\theta, \qquad y = 2\sin\theta.$$

Find the value of the integral

$$\int_{3,0}^{0,2} (3\,x^2 + 2\,xy + y^2)\,dx,$$

taken along the path C. What is the value of the integral of the given function when the path is a circle about the origin?

8. Evaluate the integral $\int_C (3\,z^2 + 7\,z + 9)\,dz$, where C is the circle $x^2 + y^2 = 3$. Is the integrand holomorphic in the region bounded by C?

9. Evaluate the integral $\int_C (2\,z^3 + 8\,z + 2)\,dz$, where C is the arc of a cycloid $y = a\,(1 - \cos\theta)$, $x = a\,(\theta - \sin\theta)$ between $(0, 0)$ and $(2\,\pi a, 0)$.

10. Show that the integrals of the functions given in Exs. 8, 9, taken about any closed curve lying in the finite region must be zero.

11. Is the integral

$$\int_{3+2\,i}^{2-3\,i} \frac{3\,z + 7}{z}\,dz$$

independent of the path of integration? What conclusion can be drawn from the answer as to the nature of the integrand?

12. Given the function: $f(z) = \dfrac{z^2 + 3z + 9}{z - 1}$. Does this function converge uniformly to its values along the circle C having the origin as a center and a radius equal to one? Does the Cauchy-Goursat theorem apply to the integral taken along C? to a circle concentric with C but lying within C?

13. Given $w = f(z) \equiv 3z^2 + 7z + 4, \quad z = \phi(\tau) \equiv \dfrac{2\tau^2 + 3}{\tau + 1}$. Is $w = f\{\phi(\tau)\}$ holomorphic for values of $|\tau| < 1$?

14. Given the function $f(z)$ defined by the relation

$$f(z) = \int_C \frac{3t^2 + 7t + 1}{t - z}\, dt,$$

where t takes complex values along the circle C of radius 2 about the origin. Compute the values of $f(z)$ and $f''(z)$ for $z = 1 + i$.

15. Given the function $f(z) = 3z^3 + 4z^2 + 7z + 2$. Find the integral of this function along the circle $x^2 + y^2 = 1$ from the point $\alpha \equiv (1, 0)$ to the point $\beta \equiv (-1, 0)$. Show that this integral taken around the complete circle is unchanged when any regular closed curve is substituted for this circle as the path of integration.

16. Given $u(x, y) = x^4 - 6x^2y^2 + y^4$. Find a function $v(x, y)$ such that $u + iv$ is an analytic function $f(z)$. Find the value of $f''(z)$ for $z = 2 + 3i$.

17. Given any rational integral function $f(z)$. Show how the value of $f(z)$ for $z = 2 + 3i$ can be found when we know the values of $f(z)$ on the circle about the origin having a radius $\rho = 4$.

18. An incompressible fluid flows over a plane with a velocity-potential

$$u = x^2 - y^2.$$

Determine a value of v such that

$$w = u + iv = f(z),$$

is holomorphic in the finite region. Find the components of the velocity and the direction of the flow at the point $z = 3 + 2i$.

CHAPTER IV

MAPPING, WITH APPLICATIONS TO ELEMENTARY FUNCTIONS

26. Conjugate functions. In the present chapter we shall discuss certain elementary functions with special reference to the correspondence between certain portions of the Z-plane and the W-plane, as determined by the relation between the function w and the independent variable z. Before taking up this general discussion, however, we shall consider the significance of u and v in the relation

$$w = f(z) = u(x, y) + iv(x, y), \tag{1}$$

where $f(z)$ is holomorphic in a given region. As we have seen, both u and v satisfy Laplace's differential equation and hence either may be considered as a potential function. They are consequently of importance in theoretical physics. Because of the relation that each function has to the other, they are called **conjugate functions.**

We can represent the functions $u(x, y)$, $v(x, y)$ by the two surfaces

$$u = u(x, y), \qquad v = v(x, y),$$

where x, y, u and x, y, v are two systems of Cartesian coördinates of points in space. These two surfaces represent then the real and the imaginary parts of the given function $w = f(z)$.

Consider, for example, the function

$$w = z^2.$$

We have then

$$w = u + iv = (x + iy)^2 = x^2 + 2\,ixy - y^2.$$

By equating the real and the imaginary parts, we get

$$u = x^2 - y^2, \qquad v = 2\,xy.$$

Each of the surfaces representing these equations is cut by any plane parallel to the XY-plane in a rectangular hyperbola. From the u-surface, we get a system of such hyperbolas having the lines $y = \pm x$ as the asymptotes. From the v-surface, we obtain a system of rectangular hyperbolas having the two axes as asymptotes.

The two systems of curves when projected upon the XY-plane appear as in Figs. 33, 34, respectively. In either case the curves are the projections of the intersections of the given surfaces with a system of equiangular hyperbolic cylinders whose generating lines are perpendicular to the XY-plane.

We have considered the two surfaces $u = u(x, y)$, $v = v(x, y)$ as related to distinct coördinate axes. Suppose now we think of them

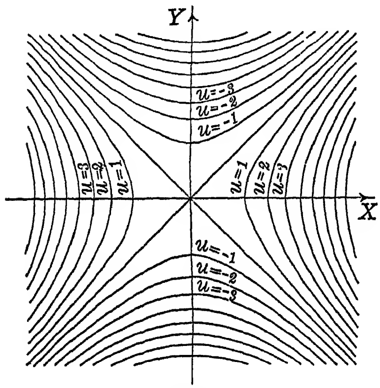

FIG. 33.

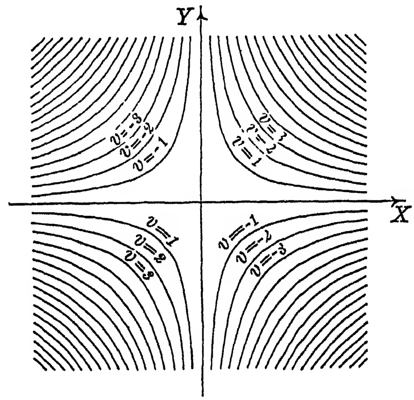

FIG. 34.

as being referred to the same system of axes. The projection upon the XY-plane of the curves of intersection reveals an important relation between the two systems of curves. It will be shown that these two systems of curves, given by

$$u = c_1, \qquad v = c_2,$$

where c_1, c_2 are constants, are orthogonal systems in the XY-plane. To do this, we make use of the slope of the curves, which is given by $\frac{dy}{dx}$. From the relation $u(x, y) = c_1$, we have for $\frac{\partial u}{\partial y} \neq 0$

$$\frac{dy}{dx} = -\frac{\frac{\partial u}{\partial x}}{\frac{\partial u}{\partial y}}.$$

In order that the curves given by $v(x, y) = c_2$ be orthogonal to the system $u(x, y) = c_1$, the slope of $v(x, y) = c_2$, at points of intersection with the curves $u(x, y) = c_1$ must be the negative reciprocal of

the slope of $u(x, y) = c_1$ at these points. The general condition of orthogonality is then

$$-\frac{\dfrac{\partial v}{\partial x}}{\dfrac{\partial v}{\partial y}} = \frac{\dfrac{\partial u}{\partial y}}{\dfrac{\partial u}{\partial x}},$$

which may be written in the form

$$\frac{\partial u}{\partial y} \cdot \frac{\partial v}{\partial y} + \frac{\partial u}{\partial x} \cdot \frac{\partial v}{\partial x} = 0.$$

This condition is satisfied by conjugate functions; for, we know that the conjugate functions $u(x, y)$ and $v(x, y)$ satisfy the Cauchy-Riemann differential equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Multiplying these two equations member by member, we have precisely the condition of orthogonality given above.

The fact that these two systems of curves are orthogonal increases the ease with which either curve may be constructed when the other is given. All we need to do is to construct a second system everywhere orthogonal to the first. When so drawn, the two sets of curves obtained in the foregoing example are as shown in Fig. 35. If we think of both systems of curves as projected back upon each of the surfaces $u = u(x, y)$, $v = v(x, y)$, we shall have upon each surface two systems of curves cutting each other at right angles.

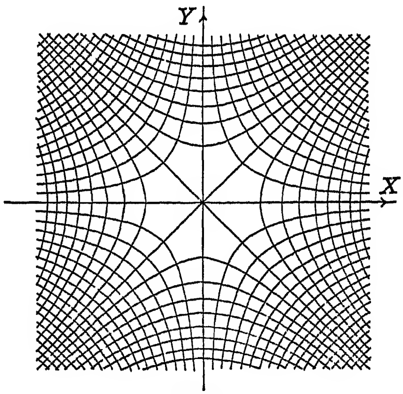

FIG. 35.

The curves cut from either surface by planes parallel to the XY-plane are called the **level lines.** The curves of the orthogonal system are called the **lines of slope,** or the curves of quickest descent. In theoretical physics other special names are employed to designate the projection of these two systems upon the XY-plane.

In an electric field or in the field of a gravitational force, a surface such that the potential is the same at all points of it is called an **equipotential surface.** Hence, any right cylinder through the lines

of level on the u-surface or the v-surface just discussed is an equipotential surface. Since one such surface may pass through each level line, we have a system of equipotential surfaces. Curves drawn perpendicular to these surfaces are called **lines of force.**

The traces of the equipotential surfaces upon the XY-plane are called **equipotential lines.** In the applications to be considered, the lines of force as well as the equipotential lines lie in a plane, which we shall take as the complex plane. In the case of the flow of an incompressible fluid or of streaming in electricity the two orthogonal systems of curves are referred to as the equipotential lines and the lines of flow respectively; while in the theory of heat they are called the isothermal lines and the lines of flow. In the case of electric currents the lines of flow are frequently called stream-lines.

We shall have frequent occasion in this chapter to return to the properties of conjugate functions.

27. Conformal mapping. The relation $w = f(z)$ gives a definite association between those points of the complex plane representing the values of z and those representing the values of w. As a matter of convenience it is usual to represent the z-points in one plane, called the Z-plane, and the w-points in another plane, called the W-plane. These two planes have a relation to each other somewhat similar to that which the two coördinate axes have in the consideration of functions of a real variable. As the point P traces any curve in the Z-plane, the corresponding point Q will trace a curve in the W-plane. We express the relation between the two curves by saying that the curve in the Z-plane is **mapped** upon the W-plane. In discussing the general properties of mapping, it is often convenient to speak of the mapping of the one plane upon the other rather than of the mapping of some particular configuration from the one plane upon the other. If $f(z)$ is multiple-valued, then to each point in the Z-plane there correspond in general several distinct points in the W-plane. In such cases it is often convenient to map the whole of the one plane upon a portion of the other.

From the discussion in Art. 21, we are able to state the conditions under which the W-plane can be mapped in a definite manner upon the Z-plane. We have seen that if the Jacobian

$$\begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\[2ex] \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}$$

does not vanish within a given region of the Z-plane, which is the case if $f'(z) \neq 0$, we can always solve the equations

$$u = \Psi_1(x, y),$$
$$v = \Psi_2(x, y) \tag{1}$$

for x, y in terms of u and v. Moreover, there is but one such solution possible. Denoting the result of this solution by

$$x = \chi_1(u, v),$$
$$y = \chi_2(u, v), \tag{2}$$

we can by means of these equations map in a definite manner the W-plane upon the Z-plane. Whether the W-plane maps upon the entire Z-plane or only upon a portion of it depends in general upon the character of the two functions χ_1, χ_2. By means of relations (1), (2), we can, however, establish a one-to-one correspondence between the points of a region of the Z-plane and those of a corresponding region of the W-plane; that is to say, if T is the region of the Z-plane under consideration and S the corresponding region of the W-plane, then to each point of T there corresponds one and only one point of S and conversely.

Ex. 1. Given $w = z^2$. Let it be required to map a given configuration from the Z-plane upon the W-plane and conversely by means of this relation.

As in Art. 25, we have

$$u = x^2 - y^2, \qquad v = 2xy.$$

We may also write

$$z = \rho(\cos\theta + i\sin\theta),$$
$$w = \rho'(\cos\theta' + i\sin\theta') = \rho^2(\cos 2\theta + i\sin 2\theta).$$

Hence, we have

$$\rho' = \rho^2, \qquad \theta' = 2\theta.$$

From the relation between θ and θ', it will be seen that a half of the Z-plane maps into the whole of the W-plane, and on the other hand a half of the W-plane maps into a quadrant of the Z-plane; for example, the upper half of the W-plane maps into the first quadrant of the Z-plane.

To map from the W-plane upon the Z-plane, suppose we put $u = c$, a constant. We obtain a rectangular hyperbola given by the equation

$$x^2 - y^2 = c.$$

Regarding c as a variable parameter, we have two systems of rectangular hyperbolas having respectively the lines $y = \pm x$ as asymptotes, according as c is positive or negative. As the upper half of the W-plane maps into the first quadrant of the Z-plane, the given lines $u = c$ map into those branches of these hyperbolas

situated in that quadrant, as represented in Fig. 37. For $v = c'$, a positive constant, we obtain in the Z-plane a system of hyperbolas orthogonal to the first systems. We see then that the upper half of the W-plane maps into the first quadrant of the Z-plane and that the orthogonal systems of lines parallel to the

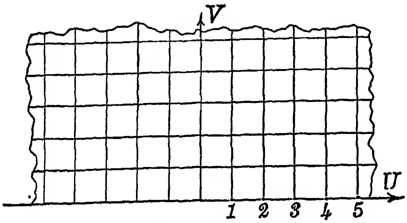

FIG. 36.

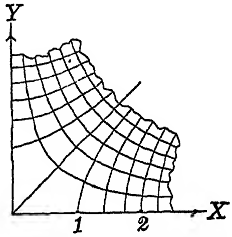

FIG. 37.

two axes of the W-plane map into orthogonal systems of rectangular hyperbolas, having respectively the two positive axes and the line $y = x$ as limiting cases.

Let us now undertake to map certain simple curves of the Z-plane upon the W-plane. We know from what has been said that a half of the Z-plane will map

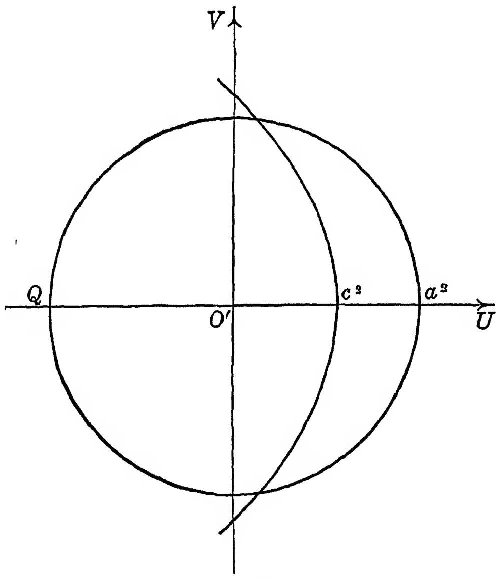

FIG. 38.

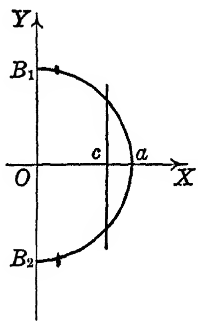

FIG. 39.

into the whole of the W-plane. Consider the line $x = 0$ of the Z-plane. We have in the W-plane

$$w = -y^2;$$

that is, for any value of y, either positive or negative, w has a negative real value. Consequently, the whole of the Y-axis maps into the negative U-axis. The

points B_1, B_2 (Fig. 39), map into the same point Q in the W-plane (Fig. 38). If z describes a semicircle of radius a as indicated, then w describes a complete circle of radius a^2 about the origin O'. If z describes the line $x = c$, then w describes a parabola cutting the U-axis at c^2; for, eliminating y between the equations

$$u = c^2 - y^2, \qquad v = 2\,cy,$$

we have

$$v^2 = 4\,c^2(c^2 - u).$$

The position of this parabola is shown in Fig. 38. In a similar manner any other curve in the Z-plane may be mapped upon the W-plane.

The given function determines an electrostatic field * in the immediate vicinity of two conducting planes at right angles to each other. In this field the equipotential surfaces are the system of hyperbolic cylinders determined by the equation $v = 2\,xy$. As a special case we have the two planes intersecting at right angles. The relation between w and z also determines the field between two coaxial rectangular hyperbolas. The system of hyperbolas $v = c$ are the lines of equipotential, while the curves of the orthogonal system $u = c$ are the lines of force.

Let us now consider the general case where one region of the complex plane is mapped upon another by means of a function $w = f(z)$ which is holomorphic for the values of z under consideration. We shall inquire into the effect of such mapping upon the angle that one curve makes with another at their point of intersection. If the magnitude of the angle is preserved even though reversed in direction the mapping is said to be **isogonal** or **conformal.** We shall now demonstrate the following proposition.

THEOREM. *The mapping of the Z-plane upon the W-plane by means of a function $w = f(z)$ is isogonal, without reversion of angles, in the neighborhood of a regular point z_0 of $f(z)$, provided $f'(z_0) \neq 0$.*

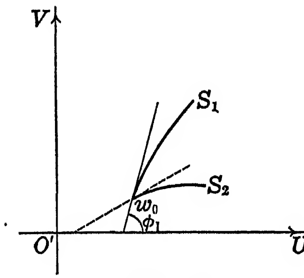

FIG. 40.

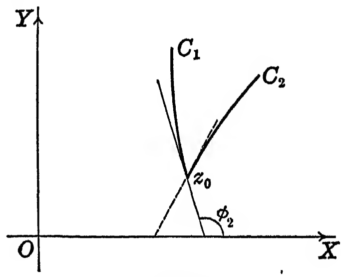

FIG. 41.

Let C_1, C_2 be any two curves in the Z-plane intersecting at z_0. Suppose C_1, C_2 map into the two curves S_1, S_2 of the W-plane intersecting in the point w_0 corresponding to z_0. We are to show that the

* See Jeans, *Electricity and Magnetism*, p. 262.

angle that C_1 makes with C_2 is the same as that which S_1 makes with S_2. The relation between w and z is given by the function

$$w = f(z).$$

Since z_0 is a regular point, the derivative of $f(z)$ exists and is defined by the relation

$$f'(z_0) = \underset{\Delta z \doteq 0}{L} \frac{\Delta w}{\Delta z}. \tag{3}$$

Since $f'(z_0)$, Δw, Δz are all complex numbers and $f'(z_0) \neq 0$, we may write

$$\begin{aligned} f'(z_0) &= \rho(\cos\theta + i\sin\theta), \qquad \rho \neq 0, \\ \Delta w &= \rho_1(\cos\theta_1 + i\sin\theta_1), \qquad \Delta z = \rho_2(\cos\theta_2 + i\sin\theta_2), \end{aligned} \tag{4}$$

where θ_1, θ_2 are taken to be the chief amplitudes of Δw, Δz, respectively.

From (3) it follows that

$$\begin{aligned} f'(z_0) &= \underset{\Delta z \doteq 0}{L} \frac{\rho_1}{\rho_2}(\cos\overline{\theta_1 - \theta_2} + i\sin\overline{\theta_1 - \theta_2}) \\ &= \underset{\Delta z \doteq 0}{L} \frac{\rho_1}{\rho_2} \underset{\Delta z \doteq 0}{L} (\cos\overline{\theta_1 - \theta_2} + i\sin\overline{\theta_1 - \theta_2}), \\ &= \underset{\Delta z \doteq 0}{L} \frac{\rho_1}{\rho_2}\left\{\cos \underset{\Delta z \doteq 0}{L} (\theta_1 - \theta_2) + i\sin \underset{\Delta z \doteq 0}{L} (\theta_1 - \theta_2)\right\}, \end{aligned} \tag{5}$$

since $\cos z$ and $\sin z$ are continuous functions.

We have therefore

$$\rho = \underset{\Delta z \doteq 0}{L} \frac{\rho_1}{\rho_2}, \quad \theta = \underset{\Delta z \doteq 0}{L} (\theta_1 - \theta_2). \tag{6}$$

Denote by ϕ_2 the angle that the tangent to the curve C_1 at z_0 makes with the positive X-axis and by ϕ_1 the angle that the tangent to the corresponding curve S_1 makes with the positive U-axis. We have then, since Δw approaches zero with Δz,

$$\underset{\Delta z \doteq 0}{L} \theta_2 = \phi_2, \qquad \underset{\Delta z = 0}{L} \theta_1 = \phi_1. \tag{7}$$

The amplitude of $f'(z_0)$ is less numerically than 2π, because of the restriction of the values of θ_1, θ_2 to the chief amplitudes of Δw, Δz. We may then write

$$\theta = \phi_1 - \phi_2; \tag{8}$$

that is to say, θ represents the angle through which the curve C_1 is turned in the process of mapping. As $f'(z_0)$ is a constant, θ is also a

constant for all curves passing through z_0; that is, every such curve is turned through the same angle θ. Hence S_1 makes the same angle with S_2 as C_1 makes with C_2. The mapping is therefore not only isogonal but the direction of the angle is not reversed.

From (6) we have

$$\rho = |f'(z_0)| = \underset{\Delta z \doteq 0}{L} \frac{\rho_1}{\rho_2},$$

which may be written in the form

$$\frac{\rho_1}{\rho_2} = \rho + \epsilon, \tag{9}$$

where ϵ vanishes with Δz. Hence, the ratio of the magnitude of an element of the resulting configuration to the magnitude of the corresponding element in the given configuration is approximately $\rho = |f'(z_0)|$, which may be called the **ratio of magnification** in the neighborhood of z_0. The approximation is closer the smaller the element. Since ρ is constant for z_0 we conclude that the similarity of infinitesimal elements is preserved.

If instead of mapping the two given curves C_1, C_2 by means of the function

$$w = f(z) = u + iv, \tag{10}$$

the mapping had been done by means of the relation

$$w_1 = f_1(z) = u - iv, \tag{11}$$

the resulting configuration would have been situated below the U-axis as shown in Fig. 42. The configurations obtained by (10) and (11) are symmetrical to each other with respect to the U-axis. We say that by this change the resulting configuration has been reflected upon the U-axis. It will be observed that in mapping by means of the function given in (11) the direction of the angle that the one curve makes with the other has been reversed in the resulting configuration. The mapping by (11) may be described as isogonal but with reversion of angles.

Mapping by means of a function which is holomorphic in a given region is but a special case of conformal mapping. One surface may be, in fact, mapped conformally upon another if we have the relation

$$ds = M\, dS,$$

where ds, dS are elements of arcs taken in any direction from corresponding points upon the two surfaces and M, the ratio of magni-

fication, depends upon the variable coördinates but is independent of the differential elements.* In the special case considered in the theorem the general factor M is replaced by $|f'(z_0)|$.

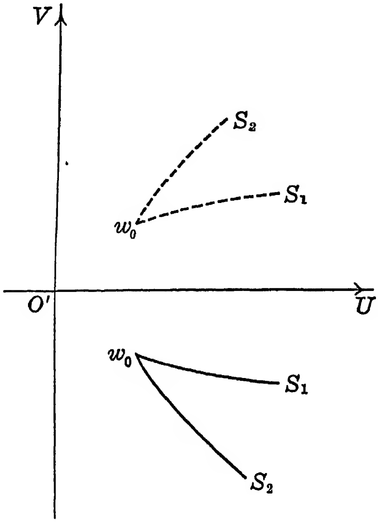

FIG. 42.

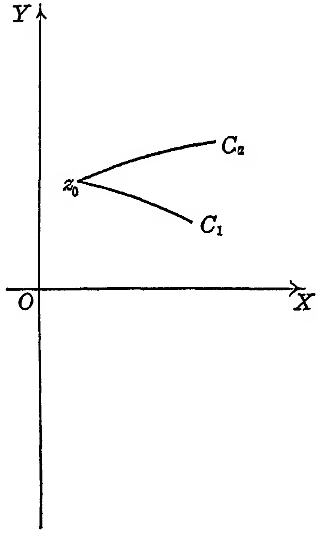

FIG. 43.

At those points of the complex plane where $f'(z) = 0$, the mapping may cease to be conformal, even if the given points are regular points of the function $w = f(z)$. For example, consider the function $w = z^2$ in the neighborhood of the point $z = 0$. Putting

$$f'(z) = \rho(\cos\theta + i\sin\theta),$$
$$z = \rho'(\cos\theta' + i\sin\theta'),$$

we have, since

$$f'(z) = D_z(z^2) = 2z,$$
$$\rho(\cos\theta + i\sin\theta) = 2\rho'(\cos\theta' + i\sin\theta').$$

But as $\rho = \rho' = 0$ for $z = 0$, this relation has no significance. As a matter of fact, as we have already seen (Ex. 1), any two curves intersecting in the origin at a given angle map by means of the function $w = z^2$ into two curves intersecting at an angle of twice that magnitude. Consequently, it can not be asserted that the mapping

* See Scheffers, *Anwendung der Differential und Integralrechnung auf Geometrie*, Vol. II, p. 70; also Osgood, *Lehrbuch der Funktionentheorie*, 2d Ed., p. 79 et seq.

by means of the functional relation $w = z^2$ is isogonal in the neighborhood of the origin.

From what has been said concerning isogonality, it must not be inferred that the map in the W-plane is as a whole identical or even similar to the original configuration in the Z-plane. The amount of distortion that takes place depends upon the coördinates of the point. To show this, consider the value of the ratio of magnification $\rho = |f'(z)|$. We have

$$\rho = |f'(z)| = \left|\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right| = \left|\frac{\partial v}{\partial y} + i\frac{\partial v}{\partial x}\right|$$
$$= \sqrt{\left(\frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2} = \sqrt{\frac{\partial u}{\partial x}\cdot\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\cdot\frac{\partial v}{\partial x}}.$$

It is evident from this relation that ρ depends upon the variables x, y, and therefore may change with the point z. The functional relation between w and z does not, therefore, necessarily establish a similarity between finite parts of the two corresponding configurations.

The geometric significance of the derivative may be regarded as a generalization of the significance of the derivative in the calculus of real variables; for, let z_0 be any given point in the Z-plane and w_0 its image in the W-plane. As z passes through the point z_0 in any direction w passes through w_0 in a corresponding direction. The modulus ρ of $f'(z_0)$ gives the limit of the ratio of the absolute value of the change that takes place in w to the absolute value of the change that takes place in z; that is, ρ measures the magnification about the point w_0 of the infinitesimal elements of the configuration in the W-plane relative to the corresponding infinitesimal elements of the Z-plane. This change corresponds in the calculus of a real variable to the change in the ordinate y as x varies, determining in that case the slope of the tangent to the curve. On the other hand, the amplitude θ of $f'(z_0)$ gives the amount of rotation between corresponding elements of the two planes. Both ρ and θ may change with z since $f'(z)$ is in general a function of z.

Further interpretations of the derivatives are frequently made in solving physical problems. If we let z move along a definite curve, then w likewise moves along some curve in the W-plane. From the relation between w and z we have

$$dw = f'(z)\,dz. \tag{12}$$

Considering the time t in which the motion takes place as the common variable in terms of which the changes of w and z are expressed, we may replace the differentials in (12) by time derivatives and write

$$D_t w = f'(z)\, D_t z, \tag{13}$$

whence

$$|D_t w| = |f'(z)| \cdot |D_t z|.$$

The derivatives $D_t w$, $D_t z$ represent the velocities both as to magnitude and direction with which the points w and z move along their respective curves. The speeds with which these motions take place are given by $|D_t w|$, $|D_t z|$ respectively. The derivative

$$f'(z) = \rho(\cos\theta + i\sin\theta) \tag{14}$$

then gives the ratio of the two velocities, while

$$\rho = |f'(z)| \tag{15}$$

gives the ratio of the speed of w to that of z.

By means of the second time derivatives of w and z the acceleration of the moving point may be determined at any instant. Differentiating (13) we have

$$D_t^2 w = f''(z) \cdot (D_t z)^2 + f'(z) \cdot D_t^2 z. \tag{16}$$

The modulus of $D_t^2 w$ gives the magnitude of the acceleration and the amplitude of $D_t^2 w$ gives the direction in which the acceleration takes place. Both the magnitude and the direction of the acceleration of the w-point involve the velocity as well as the acceleration of the corresponding z-point since $D_t^2 w$ depends upon both $D_t z$ and $D_t^2 z$. The following example illustrates the questions under discussion.

Ex. 2. Given $w = z^2$. Let z start from the point i with an initial velocity of one centimeter per second and move with uniform velocity along a line parallel to the positive axis of reals. Determine the path of the corresponding w-point and the velocity, acceleration, and speed of that point at any time t.

The path of the w-point is the map upon the W-plane of the positive half of the line $y = 1$. This line maps into the upper half of the parabola $v^2 = 4u + 4$.

The velocity ν of the w-point is given by the relation

$$\nu = \frac{dw}{dt} = f'(z) \cdot \frac{dz}{dt} = 2z \cdot \frac{dz}{dt}.$$

The derivative $\frac{dz}{dt}$ is the velocity of the z-point, which in this case is constant and equal to one centimeter per second. Hence, we have

$$\nu = 2z \cdot 1 = 2z.$$

The acceleration a is given by the relation

$$a = \frac{d^2w}{dt^2} = f''(z) \cdot \left(\frac{dz}{dt}\right)^2 + f'(z)\,\frac{d^2z}{dt^2} = 2.$$

Consequently, as the z-point starts at $z = i$ and moves as given by the conditions stated in the problem, w starts at the point $w = i^2 = -1$ and moves along the upper half of the parabola, starting with an initial velocity of

$$\nu = 2z = 2i.$$

At the end of any given time, say 3 seconds, we have

$$z = z_0 = 3 + i, \qquad w_0 = z_0^2 = 8 + 6i,$$

and

$$\nu_0 = 2z_0 = 6 + 2i.$$

The acceleration of the w-point remains constantly equal to two centimeters per second per second. The acceleration in this case being a real number its

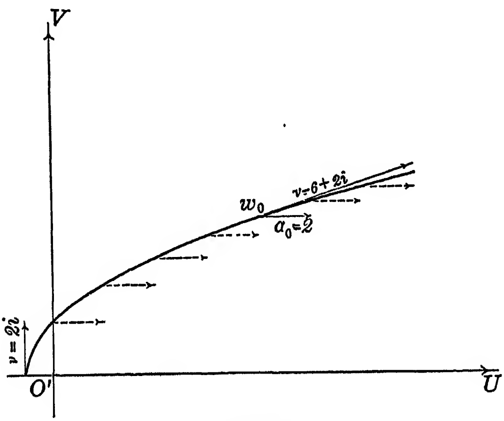

FIG. 44.

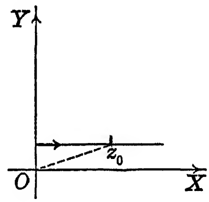

FIG. 45.

amplitude is zero and it is directed at each point parallel to the positive U-axis as shown in Fig. 44.

The direction of the velocity at any point is determined by the amplitude of z, since $\nu = 2z$. As the velocity is always measured along the tangent to the path of the moving point, it follows that the tangent to the w-curve is always parallel to the half-ray from the origin to the point z.

The speed of the w-point at any instant is

$$|D_t w| = |f'(z)| \cdot |D_t z| = 2 \cdot |z|.$$

At the end of 3 seconds the speed is then

$$2 \cdot |z_0| = 2\sqrt{9+1} = 2\sqrt{10}\,\frac{\text{cm}}{\text{sec}}.$$

28. The function $w = z^n$. In a previous article we have discussed a special case of the general function $w = z^n$, namely, the case where $n = 2$. There are some additional properties of the general case that will now be considered.

The function $w = z^2$ is the simplest case that we have of a general class of functions known as **linear automorphic functions.** By such a function we mean one that remains unchanged when the independent variable is replaced by any one of a definite set of its linear substitutions such that the set forms a group.* In this case the function remains unchanged under the linear substitutions consisting of $z = -z'$ and the identical substitution $z = z'$.

It may be shown that the general function $w = z^n$ is likewise an automorphic function. To find the particular linear transformations that leave the function unchanged, we make use of the number

$$\omega = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}.$$

This number is one of the n^{th} roots of unity; for, by De Moivre's theorem, we have

$$\omega^n = \left(\cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}\right)^n = \cos 2\pi + i\sin 2\pi = 1.$$

The other n^{th} roots, aside from unity itself, are

$$\omega^2,\ \omega^3,\ \ldots\ \omega^{n-1}.$$

Moreover, since

$$(\omega^k)^n = 1,\ (k = 1,\ 2,\ 3,\ \ldots,\ n-1),$$

we may write

$$(\omega^k \cdot z)^n = z^n.$$

Hence, if $\omega^k z'$, where $k = 0, 1, 2, 3, \ldots, n-1$, is substituted for z in the function $w = z^n$, the function remains unchanged. The substitutions $z = \omega^k z'$ form a group of linear substitutions and hence $w = z^n$ is therefore a linear automorphic function.

In the discussion of the function $w = z^2$, attention was called to the fact that a half of the Z-plane maps into the whole of the W-plane. Let us now consider the general case, where $w = z^n$. We have

$$z = \rho(\cos\theta + i\sin\theta),$$
$$w = \rho'(\cos\theta' + i\sin\theta').$$

* See Fricke, *Encyklopädie d. Math. Wiss.*, Vol. II_2, p. 351.

From these relations we obtain

$$\rho' = \rho^n, \quad \theta' = n\theta.$$

Here, as in the special case where $n = 2$, a circle about the origin in the Z-plane maps into a circle in the W-plane, and a straight line through the origin becomes a straight line through the origin in the W-plane. From the relation between θ and θ', it will be seen that $\left(\frac{1}{n}\right)^{\text{th}}$ of the circle in the Z-plane maps into the whole of the circle in the W-plane; consequently, a sector bounded by any two half-rays drawn from the origin making an angle $\frac{2\pi}{n}$ radians with each other maps into the whole of the W-plane. The values of θ' corresponding to the chief amplitude of w belong to the interval

$$-\pi < \theta' \leqq \pi.$$

The sector bounded by OR_1, OR_2 (Fig. 47), making respectively the angles $\frac{\pi}{n}$, $-\frac{\pi}{n}$ with the positive X-axis, maps in a continuous, single-

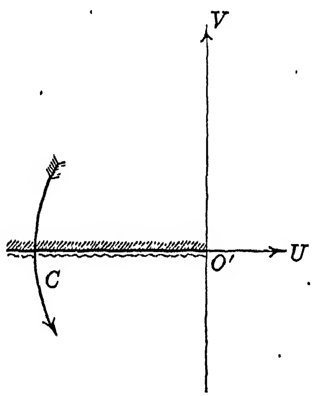

FIG. 46.

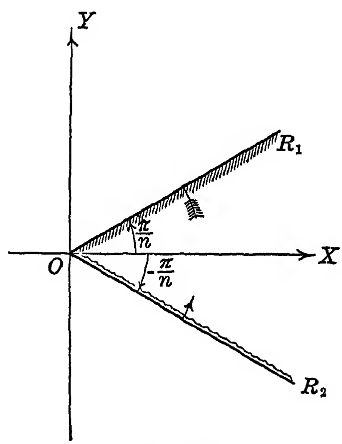

FIG. 47.

valued manner upon the entire W-plane. The lower bank of the line OR_1 goes over into the upper bank of the negative axis of reals of the W-plane. The upper bank of the line OR_2 maps into the lower bank of the negative axis of reals of the W-plane as shown in Figs. 46 and 47.

Any portion of the Z-plane which maps into the entire W-plane is called a **fundamental region** of the given function. Thus the region R_1OR_2 is a fundamental region of the function $w = z^n$. Likewise any sector bounded by two half-rays from the origin and making an angle of $\frac{2\pi}{n}$ with each other may be taken as a fundamental region. In the case of linear automorphic functions, any substitution that leaves the function unchanged maps any fundamental region of the function into another region that does not overlap the first. The second region may be taken likewise as a fundamental region of the given function.

It is to be observed that while the mapping of the Z-plane upon the W-plane by means of the given function is continuous and single-valued, it does not follow that the mapping of the W-plane back upon the fundamental region of the Z-plane is continuous. As a matter of fact, such a mapping is single-valued but not continuous. In other words, not every continuous curve in the W-plane maps into a continuous curve in the fundamental region of the Z-plane. Suppose, for example, that we have a curve in the W-plane crossing the negative axis of reals. As we have already seen, the upper bank of the negative U-axis maps into the lower bank of the line OR_1; while, on the other hand, the lower bank of the negative U-axis maps into the upper bank of the line OR_2. Consequently, the curve C (Fig. 46), which is a continuous curve in the W-plane, becomes a discontinuous curve in the Z-plane. We can say, however, that the mapping from the W-plane to the fundamental region of the Z-plane is a single-valued process; for, to every point of the W-plane there is one and only one corresponding point in the fundamental region of the Z-plane.

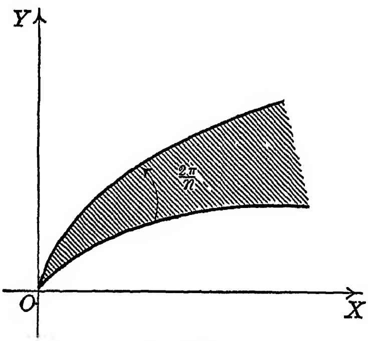

FIG. 48.

A fundamental region of the function $w = z^n$ need not be bounded by straight lines. It serves our purpose equally well to take any continuous curve proceeding from the origin and to revolve it through the angle $\frac{2\pi}{n}$, as indicated in Fig. 48. The boundary lines of the

sector extend indefinitely from the origin in the case of the function under discussion.

It must not be assumed that the fundamental region of all automorphic functions can be obtained in this manner. The function here considered is a special case. A more general case will be discussed in a subsequent chapter, in connection with doubly periodic functions.

By means of the fundamental operations of multiplication and addition applied to complex constants and functions of the type $w = z^n$, where n is integral, we obtain the rational integral function

$$f(z) = \alpha_0 z^n + \alpha_1 z^{n-1} + \alpha_2 z^{n-2} + \cdots + \alpha_n, \quad \alpha_0 \neq 0.$$

It is of interest in this connection to point out some of the applications * that may be made in theoretical physics of the transformation $w = z^n$.

For the special case where $n = -1$, we have

$$w = \frac{1}{z},$$

or

$$u + iv = \frac{1}{x + iy},$$

whence

$$u = \frac{x}{x^2 + y^2}, \qquad v = \frac{-y}{x^2 + y^2}, \tag{1}$$

Corresponding to the lines $u = c$ we have

$$\frac{x}{x^2 + y^2} = c,$$

or

$$c(x^2 + y^2) - x = 0.$$

This equation is represented by a system of circles having their centers on the X-axis and all passing through the origin. For the orthogonal system, we have

$$\frac{-y}{x^2 + y^2} = c,$$

or

$$c(x^2 + y^2) + y = 0,$$

* See Jeans, *Electricity and Magnetism*, p. 261 et seq; Lamb, *Hydrodynamics*, 3[d] Ed., p. 66.

which is represented by a system of circles having their centers on the Y-axis. These two systems of circles are shown in Fig. 49.

When a fluid flows over a plane surface, any point P from which the fluid flows out in all directions in a uniform manner is called a **source.** By the strength of the source is understood the total flow in a unit of time across a small closed curve about the source, that is,

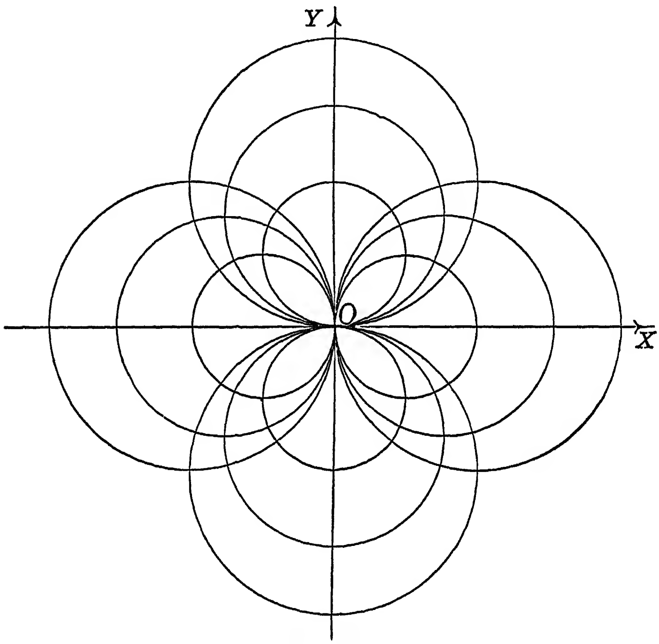

FIG. 49.

the time rate of the supply of the fluid through the source. A negative source is called a **sink.** Let P' be a sink and suppose it to be of equal strength with the source P. Denote this common strength by m. If we now think of P as approaching P' in such a manner that the product $m \cdot \overline{PP'}$ remains constant, say equal to n, then we say that we have a **plane doublet** * of strength n. The velocity-potential due to the doublet is given by

$$\phi(x, y) = \frac{-nx}{x^2 + y^2},$$

* See Pierce, *Newtonian Potential Function*, p. 434, et seq.

while the lines of flow are given by

$$\psi(x, y) = \frac{ny}{x^2 + y^2} = c.$$

By comparing these functions with the conjugate functions $u(x, y)$, $v(x, y)$ given in the equations (1), it will be seen that u, v determine the velocity-potential and the lines of flow respectively of a plane doublet at the origin whose strength n is -1. The X-axis is the axis of the doublet. The lines of flow $v = c$ are, as we have seen, the system of coaxial circles having their centers on the Y-axis, while the system of coaxial circles having their centers on the X-axis are the lines of equal velocity-potential.

Writing the given function $w = z^n$ in the form

$$u + iv = \rho^n(\cos\theta + i\sin\theta)^n;$$

we have, upon equating the real parts and likewise the imaginary parts,

$$u = \rho^n \cos n\theta, \qquad v = \rho^n \sin n\theta.$$

For $n = 1$ the first of these equations gives, for the flow of an incompressible fluid, a system of equipotential curves parallel to the Y-axis, and the second gives as the corresponding lines of flow a system of lines parallel to the X-axis.

It has been pointed out that when $n = 2$, then the equation $u = c$ gives a system of rectangular hyperbolas having the axes of coördinates as their principal axes. In the applications to the flow of an incompressible fluid, these curves are the lines of equal velocity-potential of an irrotational fluid having constant density and a steady flow. The curves $v = c$, that is the lines of flow are likewise a system of rectangular hyperbolas, having in this case the axes of coördinates as asymptotes. The lines $\theta = 0$, $\theta = \frac{\pi}{2}$ are parts of the same line of flow corresponding to $v = 0$; hence, we may take the positive parts of the coördinates axes as fixed boundaries, and thus obtain a case of irrotational fluid motion in an angle between two perpendicular walls (see Fig 37).

By the proper selection of the value of n, we may so change the conditions as to represent an irrotational motion of the fluid between two rigid walls making a given angle ϕ with each other. The lines of flow are given by the equation

$$\rho^n \sin n\theta = c,$$

where the lines $\theta = 0$, $\theta = \frac{\pi}{n}$ are parts of the same line of flow, namely the one by putting $c = 0$. If now we so select n that the angle $\frac{\pi}{n}$ shall be the required angle ϕ, that is if $n = \frac{\pi}{\phi}$, we get as the required equations of the equipotential curves and lines of flow, respectively, the following:

$$u = \rho^{\frac{\pi}{\phi}} \cos \frac{\pi\theta}{\phi}, \qquad v = \rho^{\frac{\pi}{\phi}} \sin \frac{\pi\theta}{\phi}.$$

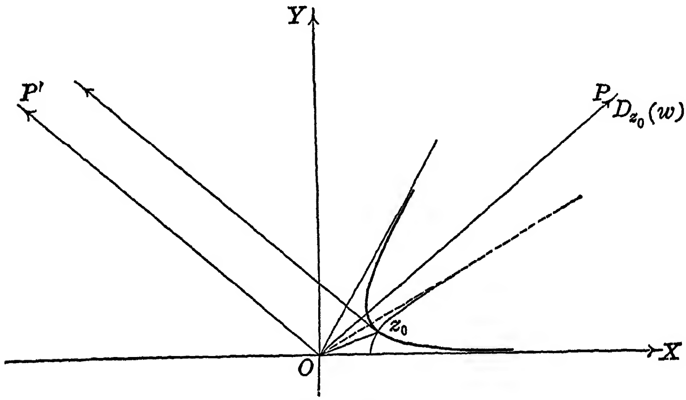

FIG. 50.

Ex. Find the value of n such that the function $w = z^n$ shall determine an irrotational motion of a liquid between two walls making an angle of 60° with each other. Find the velocity-potential and direction of the flow at the point $z_0 = 2(\cos 20° + i \sin 20°)$, assuming the flow to be steady. Trace the equipotential curve through the given point.

We have

$$n = \frac{\pi}{\frac{\pi}{3}} = 3,$$

and therefore $\qquad w = z^3 = u + iv,$

whence $\qquad u = x^3 - 3xy^2, \qquad v = 3x^2y - y^3.$

The velocity-potential at the given point is

$$u = \rho^3 \cos 3\theta$$
$$= 2^3 \cos 3 \cdot 20° = 4.$$

The equipotential curve is then given by the equation

$$x^3 - 3xy^2 = 4.$$

From the definition of velocity-potential, we know that the components of the velocity in the direction of the X-axis and Y-axis are $-\frac{\partial u}{\partial x}$, $-\frac{\partial u}{\partial y}$, respectively. The velocity ν is given by

$$\nu = -\frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}. \tag{8}$$

We can determine graphically the velocity by means of the derivative $D_z w$. As we have seen, Art. 21,

$$D_z w = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}. \tag{9}$$

By comparison of (8) and (9), it will be seen that the point P' representing the velocity at z_0 is the reflection upon the axis of imaginaries of the point P representing $D_z w$. We have

$$D_{z_0} w = 3\, z_0^2 = 12\,(\cos 40° + i \sin 40°).$$

The point P is found by laying off on a line making an angle of 40° with the axis of reals the distance $OP = 12$. By reflection upon the Y-axis the point P' is found, which represents the velocity at z_0. Drawing from z_0 a line parallel to OP', we have the direction of the flow at z_0. This flow is in the direction of the normal to the equipotential curve through z_0 as we should expect.

It should also be observed that if the point z_0 is allowed to move along a curve of flow

$$v \equiv 3\, x^2 y - y^3 = k,$$

the point P' varies in such a manner as to describe the hodograph* of the motion of z_0, and thus the velocity of P' determines the magnitude and direction of the acceleration of z_0.

The relation between the velocity in the Z-plane and the derivative $D_z w$, as brought out in the above exercise, is perfectly general and may be applied to any case, so long as the function $w = f(z)$ is holomorphic in the region under consideration. It should also be noted in this connection that the velocity of a point moving along the line $v = c$ in the W-plane is always the negative of the square of the speed of the corresponding point moving along the curve of flow in the Z-plane; for, denoting by ν_w, ν_z, the velocities in the W-plane and Z-plane respectively, we have by equation (13), Art. 27,

$$\begin{aligned}\nu_w = D_z w \cdot \nu_z &= \left(\frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}\right)\left(-\frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}\right)\\ &= -\left(\frac{\partial u}{\partial x}\right)^2 - \left(\frac{\partial u}{\partial y}\right)^2\\ &= -\,|\,\nu_z\,|^2 = -\,r^2,\end{aligned}$$

where r is the speed along the curve of flow in the Z-plane.

* See Ziwet, *Theoretical Mechanics*, Part I, p. 80.

29. Definition and properties of e^z. We shall now define the exponential function e^z, where z is a complex variable. The definition and properties of the function e^x, where x is a real variable, can not be assumed to hold when the variable is complex. The function e^z should be so defined, however, as to include, as a special case, the function e^x. It will assist us in formulating this definition to recall the definition and some of the general properties of e^x. The number e is defined by the limit

$$e = \underset{n=\infty}{L}\left(1+\frac{1}{n}\right)^n = 2.7182818\ \ldots,$$

where n takes all positive real values. Likewise, the exponential function e^x is defined as the limit $\underset{n=\infty}{L}\left(1+\frac{x}{n}\right)^n$.

This function obeys the general law,

$$f(x_1)\cdot f(x_2) = f(x_1+x_2). \tag{1}$$

Differentiating the function $f(x) = e^x$, we have

$$D_x f(x) = \frac{d(e^x)}{dx} = e^x. \tag{2}$$

We shall define the exponential function of a complex variable by the relation,

$$e^z = \underset{n=\infty}{L}\left(1+\frac{z}{n}\right)^n, \tag{3}$$

where n is a positive integer. We shall show that this limit exists for all finite values of z and that the function thus defined has the general properties expressed for e^x in (1) and (2).

To show the existence of this limit, we proceed as follows. We write

$$\begin{aligned} 1+\frac{z}{n} &= 1+\frac{x+iy}{n} \\ &= 1+\frac{x}{n}+i\frac{y}{n}. \end{aligned} \tag{4}$$

Putting

$$1+\frac{x}{n} = \rho\cos\theta, \qquad \frac{y}{n} = \rho\sin\theta, \tag{5}$$

we get

$$\begin{aligned} \left(1+\frac{z}{n}\right)^n &= [\rho(\cos\theta+i\sin\theta)]^n \\ &= \rho^n(\cos n\theta + i\sin n\theta). \end{aligned} \tag{6}$$

Since n can be taken so large that $1+\dfrac{x}{n}$, and therefore $\cos\theta$, is always positive, from (5) we obtain θ as the principal value of $\arctan\dfrac{y}{n+x}$, and

$$\rho=\left[\left(1+\frac{x}{n}\right)^2+\frac{y^2}{n^2}\right]^{\frac{1}{2}},$$

whence

$$\rho^n=\left[\left(1+\frac{x}{n}\right)^2+\frac{y^2}{n^2}\right]^{\frac{n}{2}}$$

$$=\left(1+\frac{x}{n}\right)^n\left[1+\frac{y^2}{(n+x)^2}\right]^{\frac{n}{2}}. \tag{7}$$

The limit in (3) may then be written

$$\underset{n=\infty}{L}\left(1+\frac{z}{n}\right)^n=\underset{n=\infty}{L}\left\{\left(1+\frac{x}{n}\right)^n\left[1+\frac{y^2}{(n+x)^2}\right]^{\frac{n}{2}}\left(\cos n\arctan\frac{y}{n+x}+i\sin n\arctan\frac{y}{n+x}\right)\right\}$$

$$=\underset{n=\infty}{L}\left(1+\frac{x}{n}\right)^n\cdot\underset{n=\infty}{L}\left(1+\frac{y^2}{(n+x)^2}\right)^{\frac{n}{2}}\cdot\left[\underset{n=\infty}{L}\cos n\arctan\frac{y}{n+x}+i\underset{n=\infty}{L}\sin n\arctan\frac{y}{n+x}\right] \tag{8}$$

provided each of these limits exist. These limits, however, do exist and can be readily evaluated. We have from functions of a real variable

$$\underset{n=\infty}{L}\left(1+\frac{x}{n}\right)^n=e^x.$$

For $y=0$, the second limit in (8) is one; for $y\neq 0$, we have

$$\underset{n=\infty}{L}\left[1+\frac{y^2}{(n+x)^2}\right]^{\frac{n}{2}}=\left\{\underset{n=\infty}{L}\left[1+\frac{y^2}{(n+x)^2}\right]^n\right\}^{\frac{1}{2}}. \tag{9}$$

However, we have

$$\underset{n=\infty}{L}\left[1+\frac{y^2}{(n+x)^2}\right]^n=\left\{\underset{n=\infty}{L}\left[1+\frac{1}{\frac{(n+x)^2}{y^2}}\right]^{\frac{(n+x)^2}{y^2}}\right\}^{\underset{n=\infty}{L}\, n\cdot\frac{y^2}{(n+x)^2}}=e^0.$$

Hence, from (9) we get

$$\underset{n=\infty}{L}\left[1+\frac{y^2}{(n+x)^2}\right]^{\frac{n}{2}} = (e^0)^{\frac{1}{2}} = 1.$$

Since the cosine is a continuous function, we may write the third limit in (8) in the form

$$\underset{n=\infty}{L} \cos n \arctan \frac{y}{n+x} = \cos \underset{n=\infty}{L}\; n \arctan \frac{y}{n+x}$$

$$= \cos \underset{n=\infty}{L} \frac{ny}{n+x} \cdot \frac{\arctan \dfrac{y}{n+x}}{\dfrac{y}{n+x}} = \cos y. \tag{10}$$

Similarly, we have for the final limit in (8)

$$\underset{n=\infty}{L} \sin n \arctan \frac{y}{n+x} = \sin y. \tag{11}$$

Hence, substituting these results in (8) we obtain

$$e^z = e^x(\cos y + i \sin y). \tag{12}$$

This result not only shows that the limit given in (3) exists for all finite values of z, but it gives a very convenient form of the definition of the exponential function e^z.

From this form of the definition, we can deduce a convenient method for writing the complex number

$$z = \rho(\cos\theta + i \sin\theta); \tag{13}$$

for, putting $x = 0$ in (12), we have

$$e^{iy} = \cos y + i \sin y,$$

or writing this result in the usual form, we have

$$e^{i\theta} = \cos\theta + i \sin\theta.$$

Hence we may write (13) in the form

$$z = \rho e^{i\theta}, \tag{14}$$

a form of expression that is often convenient.

The function e^z is uniquely determined; for, we have from (12)

$$e^z = u + iv = e^x \cos y + ie^x \sin y,$$

whence

$$u = e^x \cos y, \qquad v = e^x \sin y. \tag{15}$$

From these equations it follows that the conditions that $f(z) = e^z$ is an analytic function are satisfied; for, we have for all finite values of x, y

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Therefore, e^z is holomorphic in the finite region of the complex plane and consequently is an analytic function. Hence, in the finite region e^z is continuous and has a continuous derivative. Moreover, e^z is a single-valued function of z, and e^x appears as a special case.

From (12) it can be shown that the general properties of the exponential function of a real variable may be extended to the case where the variable is complex. For example we may deduce as follows the general law expressed in (1), namely,

$$e^{z_1} \cdot e^{z_2} = e^{z_1+z_2}.$$

Substituting the values of e^{z_1}, e^{z_2}, as defined by (12), we have

$$\begin{aligned} e^{z_1} \cdot e^{z_2} &= e^{x_1}(\cos y_1 + i \sin y_1)\, e^{x_2}(\cos y_2 + i \sin y_2) \\ &= e^{x_1+x_2}\{\cos(y_1 + y_2) + i \sin(y_1 + y_2)\} \\ &= e^{z_1+z_2}. \end{aligned}$$

The law of differentiation stated in (2) holds also where the variable is complex. Remembering that

$$D_z w = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x},$$

we have from (12)

$$\begin{aligned} D_z e^z &= e^x \cos y + i e^x \sin y \\ &= e^x(\cos y + i \sin y) \\ &= e^z. \end{aligned}$$

It is to be observed that e^z is a **periodic function**; that is, the function remains invariant when z is replaced by z plus some constant, say ω. Such a function satisfies the relation

$$f(z + \omega) = f(z).$$

The constant ω is called a **period** of the given function. A periodic function takes all of its values as the variable z takes the values in a definite region of the complex plane, known as the **region of periodicity**, and repeats those values as z varies over another equal portion of the plane. In this particular case the regions of periodicity are parallel strips bounded by lines parallel to the axis of

reals and at a distance of 2π from each other. We say then that the function has the period $2\pi i$. To show this to be the case, we may write

$$\begin{aligned} e^{z+2\pi i} &= e^{x+i(y+2\pi)} \\ &= e^x \{\cos(y+2\pi) + i\sin(y+2\pi)\} \\ &= e^x \{\cos y + i \sin y\} \\ &= e^z. \end{aligned}$$

Instead of $2\pi i$, we could have taken equally well any multiple of it as a period. It would not have answered the purpose to have taken a fractional part of $2\pi i$ nor any number less than $2\pi i$; that is $2\pi i$ is the smallest constant that answers the purpose. We indicate this fact by calling $2\pi i$ the **primitive period** of the function. When, as in this case, all of the periods are multiples of a single primitive period, the function is called a **simply periodic function.**

Let us now undertake to map the Z-plane upon the W-plane, and conversely, by means of the relation $w = e^z$. Since the given function is holomorphic having a derivative different from zero for finite values of z, it follows that the mapping is conformal in the

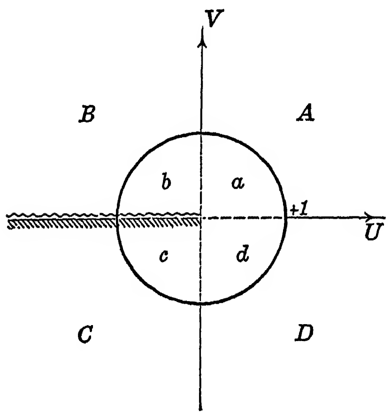

FIG. 51.

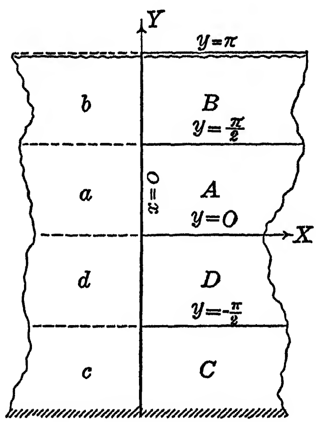

FIG. 52.

finite region; that is, in this region the similarity of infinitesimal elements is preserved. We shall make use of the fact that the function is periodic, having the period $2\pi i$. If we draw through the points πi and $-\pi i$ in the Z-plane two lines parallel to the X-axis, the region bounded by these lines maps into the entire W-plane and therefore may be taken as the fundamental region. In

mapping we shall consider the boundary line $y = \pi$, but not the boundary line $y = -\pi$, as belonging to the fundamental region. What is said of this region may be said of any one of the regions bounded by the lines

$$y = (2k+1)\pi, \quad y = (2k-1)\pi, \quad k = \cdots, -2, -1, 0, 1, 2, \cdots.$$

The line $u = 0$ is the map of certain lines parallel to the X-axis. To show this, put $u = 0$ in (15) and thus obtain

$$0 = e^x \cos y, \tag{16}$$

whence for finite values of x, we have $u = 0$ as the map of the lines

$$y = (2k+1)\frac{\pi}{2}, \quad k = \cdots, -2, -1, 0, +1, +2, \cdots. \tag{17}$$

Within the fundamental region $-\pi < y \leqq \pi$, we have the lines corresponding to $k = 0$ and $k = -1$; and hence the line $u = 0$, that is the V-axis, is the map of two lines of the Z-plane lying within this region, namely:

$$y = \frac{\pi}{2} \quad \text{and} \quad y = -\frac{\pi}{2}. \tag{18}$$

For $y = \frac{\pi}{2}$ we have from (15) $v = e^x$, and for $y = -\frac{\pi}{2}$ we have $v = -e^x$; so that the positive V-axis is the map of the line $y = \frac{\pi}{2}$, while the negative V-axis is the map of the line $y = -\frac{\pi}{2}$. Since for $x > 0$, $y = \pm\frac{\pi}{2}$ we have $|e^z| = e^x > 1$, it follows that the portion of the positive V-axis exterior to the unit circle about the origin is the map of the positive half of the line $y = \frac{\pi}{2}$, and the portion of the negative V-axis exterior to the unit circle is the map of the positive half of the line $y = -\frac{\pi}{2}$. Likewise, since for $x < 0$, $y = \pm\frac{\pi}{2}$ we have $|e^z| = e^x < 1$, it follows that the negative halves of the lines $y = \frac{\pi}{2}$, $y = -\frac{\pi}{2}$ map respectively into the portions of the positive and the negative V-axis lying within the unit circle.

In a similar manner we have from (15), for $v = 0$,

$$0 = e^x \sin y, \tag{19}$$

and hence the line $v = 0$, that is the U-axis, is the map of the lines

$$y = k\pi, \quad k = \cdots, -2, -1, 0, +1, +2, \cdots. \tag{20}$$

For the fundamental region $-\pi < y \leqq \pi$, we have $k = 0$, 1, and consequently

$$y = 0, \pi. \tag{21}$$

For these values of y, we obtain

$$u = e^x, \quad u = -e^x, \tag{22}$$

respectively. Hence the positive U-axis is the map of the X-axis, and the negative U-axis is the map of the line $y = \pi$.

The positive halves of the lines $y = 0$ and $y = \pi$ map into that portion of the U-axis exterior to the unit circle, while the negative halves of the same lines map into that portion of the U-axis within the unit circle.

Any line parallel to the X-axis maps into a half-ray in the W-plane proceeding from the origin, Fig. 53. This may be shown as follows.

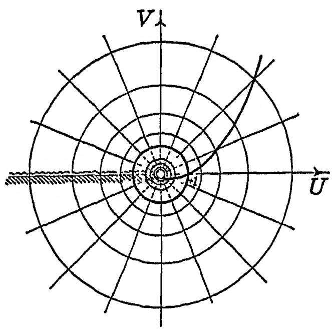

FIG. 53.

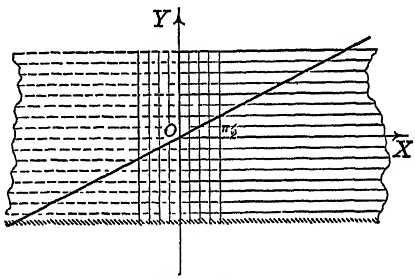

FIG. 54.

Eliminating x from equations (15) by multiplying the first of these equations by $\sin y$ and the second by $\cos y$ and subtracting, we have

$$u \sin y - v \cos y = 0.$$

For constant values of y, this equation gives straight lines of the form

$$v = mu,$$

where $m \equiv \tan y$. Since e^x is positive for all finite values of x, it follows from (15) that any line $y = c$ maps into a half-ray from the origin taken along the line $v = mu$; the portion of this half-ray

interior to the unit circle corresponds to negative values of x, while the portion exterior to the unit circle corresponds to positive values of x. If successive values of y differ by equal amounts, then the corresponding half-rays in the W-plane will make equal angles with each other.

The map of a line parallel to the Y-axis may be easily obtained as follows. Eliminating y from the equations

$$u = e^x \cos y, \quad v = e^x \sin y$$

by squaring and adding, we have

$$\begin{aligned} u^2 + v^2 &= e^{2x} (\cos^2 y + \sin^2 y) \\ &= e^{2x}. \end{aligned}$$

For any constant value of x, we have then a circle in the W-plane about the origin as a center.

For $$x = 0,$$
we have $$u^2 + v^2 = 1;$$

that is, the Y-axis maps into the unit circle about the origin in the W-plane. For $x = c > 0$, the map in the W-plane is a circle exterior to the unit circle; and for $x = c < 0$, the map is a circle lying within the unit circle. From what has been said, it will now be seen that the regions a, b, c, d, A, B, C, D, Fig. 52, map respectively into the regions a, b, c, d, A, B, C, D, Fig. 51, the lower bank of the line $y = \pi$, and the upper bank of the line $y = -\pi$ mapping respectively into the upper and the lower banks of the negative U-axis.

Any line $$y = mx$$

passing through the origin, other than the axes of coördinates, maps into a curve in the W-plane that cuts the half-rays from the origin at a constant angle; that is, it maps into a logarithmic spiral about the origin. Since the point $x = 0$, $y = 0$ maps into the point $u = 1$, $v = 0$, all of these logarithmic spirals pass through the point $u = 1$, $v = 0$.

As we have seen the whole of the W-plane can be mapped into any one of a number of strips parallel to the axis of reals in the Z-plane. Suppose we map it into the fundamental region lying between the lines $y = \pi$, $y = -\pi$. Consider the line $u = c$, $c > 0$. For this value of u, we have

$$c = e^x \cos y,$$

or $$y = \operatorname{arc\,sec} \frac{e^x}{c}. \tag{23}$$

If we put $e^x = c$, that is $x = \log c$, we have

$$y = \text{arc sec } 1 = 0.$$

The curve whose equation is (23) then cuts the X-axis at the point $x = \log c$. As x increases y increases and approaches asymptotically the line $y = \frac{\pi}{2}$. The sign of y is determined by the sign of v in the relation

$$v = e^x \sin y. \tag{24}$$

Since e^x is always positive, $\sin y$ and therefore y is positive or negative according as v is positive or negative. The curve is therefore

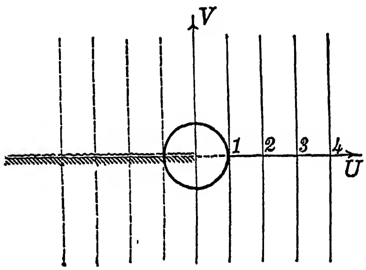

FIG. 55.

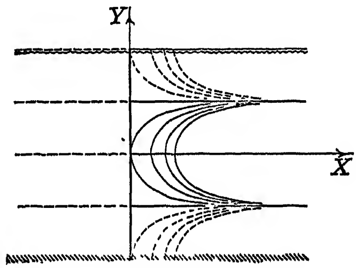

FIG. 56.

symmetrical with respect to the X-axis and is situated as is indicated in Fig. 56. As c is assigned different values, the point where the curve crosses the X-axis changes. For $c > 1$, the curve cuts the X-axis to the right of the point $x = 0$; for $c = 1$, it crosses at the origin; for $0 < c < 1$ it crosses to the left of the origin.

It will be remembered that the negative half of the U-axis maps into the lines $y = \pi$, $y = -\pi$. That portion of the line $u = c$, where $c < 0$, lying above the U-axis will, as we have seen, map into the portion of the fundamental region lying between the lines $y = \frac{\pi}{2}$ and $y = \pi$. Moreover, since c is now negative it follows that y decreases as x increases. The form of the curve is indicated in Fig. 56. It is asymptotic to the line $y = \frac{\pi}{2}$. In the same way it follows that the portion of $u = c$ lying below the U-axis maps into a curve beginning on the line $y = -\pi$ and becoming asymptotic to the straight line $y = -\frac{\pi}{2}$, since the ordinate increases with x. The results of map-

ping the W-plane upon the fundamental region $-\pi < y \leqq \pi$ are of course repeated in any other strip bounded by the lines $y = (2k + 1)\pi$, $y = (2k - 1)\pi$.

It is of interest in this connection to observe the form of the surface $\zeta = u(x, y)$. The curves just obtained by mapping upon the Z-plane the lines $u = c$ are the curves of intersection of this surface by the plane $\zeta = c$. A general notion of the form of the surface is obtained by noting the manner in which u changes as x increases along certain lines parallel to the X-axis. Take for this purpose the lines

$$y = -\pi, \quad -\frac{\pi}{2}, \quad 0, \quad \frac{\pi}{2}, \quad \pi.$$

From (15) it will be seen that along these lines, we have

$$u = -e^x, \quad 0, \quad e^x, \quad 0, \quad -e^x.$$

As x decreases without limit through negative values, each of these values of u approaches zero. However, as x increases the value of u remains zero along the lines $y = -\frac{\pi}{2}, \frac{\pi}{2}$, but increases without limit along the line $y = 0$, and decreases without limit along the lines $y = -\pi, \pi$. Hence, we have a surface that is flat at the extreme left and towards the right has ridges and valleys of increasing magnitude. These ridges and valleys extend parallel to the axis of reals and their magnitude is readily determined by taking a cross-section of the surface parallel to the Y-axis.

As an illustration, let us consider the function

$$z = w + e^w.$$

As we shall see later, this function is of importance in the consideration of certain problems in mathematical physics. We shall map a given configuration from the W-plane upon the Z-plane by means of this relation. In order to do so, we must first obtain x and y in terms of u and v. Writing the given function in the form

$$\begin{aligned} x + iy &= u + iv + e^{u+iv} \\ &= u + iv + e^u \cdot e^{iv} \\ &= u + iv + e^u (\cos v + i \sin v), \end{aligned}$$

we have upon equating the real and the imaginary parts

$$x = u + e^u \cos v, \qquad y = v + e^u \sin v. \tag{25}$$

The axis $v = 0$ maps into the X-axis; for, we have in this case from the equations (25)

$$x = u + e^u, \quad y = 0,$$

and, consequently, every point on the U-axis maps into a point on the X-axis, the point $w = 0$ mapping into $z = +1$ (Fig. 57). For $v = \pi$, we have

$$x = u - e^u, \quad y = \pi.$$

As w moves along the line $v = \pi$ from $u = -\infty$ to $u = 0$, the corresponding point in the Z-plane moves along the line $y = \pi$ from $x = -\infty$ to $x = -1$. As the value of u continues to increase from $u = 0$ to $u = \infty$, the value of x passes from

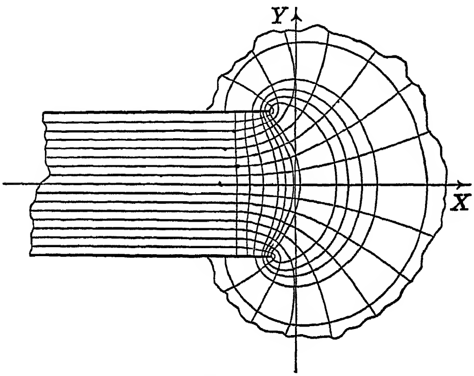

FIG. 57.

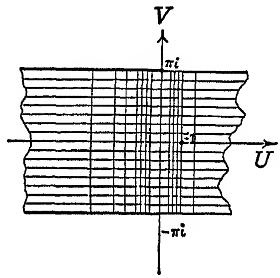

FIG. 58.

$x = -1$ back to $x = -\infty$. We say that the line $v = \pi$ maps into the line $y = \pi$ in such a manner that the line doubles back upon itself at the point $x = -1$.

In a similar manner, the line $v = -\pi$ maps into the line $y = -\pi$, bending back upon itself at the point $x = -1$.

Let us now consider the line $v = \dfrac{\pi}{2}$. For this value of v we have

$$x = u, \quad y = \frac{\pi}{2} + e^u.$$

For $u = -\infty$, we have $x = -\infty, y = \dfrac{\pi}{2}$. As u increases, y increases until u reaches the value zero, where $x = 0$, $y = \dfrac{\pi}{2} + 1$. As u increases through positive values, y continues to increase with u as indicated in the figure.

For values of v lying between $\dfrac{\pi}{2}$ and π, say for $(\pi - \epsilon)$, we have

$$x = u - e^u \cos \epsilon,$$
$$y = \pi - \epsilon + e^u \sin \epsilon.$$

From these equations, we have for $u = -\infty$,

$$x = -\infty,$$
$$y = \pi - \epsilon.$$

As u increases, both x and y increase, although y increases very slowly, until we have

$$D_u x = 1 - e^u \cos \epsilon = 0;$$

that is, until

$$e^u \cos \epsilon = 1.$$

As $e^u \cos \epsilon$ becomes greater than 1, $D_u x$ becomes negative and x decreases while y continues to increase and that more rapidly. The general form of such a curve is shown in the figure.

For values of c lying between 0 and $\frac{\pi}{2}$, the line $v = c$ maps into a curve such that y at first increases very slowly and then more rapidly as u takes on large positive values. In this case, however, x also continues to increase as u increases.

For values of v less than 0, the curves lie below the X-axis and are symmetrical as to that axis with those already obtained. The mapping of these curves therefore presents nothing new.

The given function expresses the motion of a fluid from a reservoir of indefinitely large size into a narrow, restricted channel bounded by thin parallel walls.* If the sign of w is changed, the given function represents the flow as taking place in the opposite direction. As may be shown, the velocity of the flow increases indefinitely in the neighborhood of the points $(-1, \pi)$, $(-1, -\pi)$.

30. The function $w = \log z$. We shall now define the logarithmic function and discuss some of its properties. In real variables the logarithm is frequently defined as the inverse function of the exponential. This property will be used in defining the logarithm of a complex number. The general properties of inverse functions have been discussed in another connection. To determine the inverse of the exponential function let us consider the relation

$$e^w = z. \tag{1}$$

We have, as elsewhere,

$$w = u + iv;$$
$$z = \rho(\cos \phi + i \sin \phi).$$

Equation (1) may then be written in the form

$$e^{u+iv} = e^u \cdot e^{iv} = \rho\,(\cos \phi + i \sin \phi). \tag{2}$$

Remembering that

$$e^{iv} = \cos v + i \sin v,$$

we may now write (2) in the form

$$e^u(\cos v + i \sin v) = \rho(\cos \phi + i \sin \phi).$$

Equating the real and the imaginary parts in this equation, we obtain

$$e^u \cos v = \rho \cos \phi, \quad e^u \sin v = \rho \sin \phi. \tag{3}$$

* See Lamb, *Hydrodynamics*, 3d Ed., p. 70.

Each number entering into these equations is real, and consequently we can solve the equations for u and v by the means already at our disposal. Squaring each member of these equations and adding, we have

$$(e^u)^2 (\cos^2 v + \sin^2 v) = \rho^2(\cos^2 \phi + \sin^2 \phi),$$

whence

$$(e^u)^2 = \rho^2,$$

but as e^u and ρ are always positive numbers, we may write

$$e^u = \rho. \tag{4}$$

Making use of this relation, we have from (3)

$$\cos v = \cos \phi,$$
$$\sin v = \sin \phi,$$

whence

$$v = \phi. \tag{5}$$

Since u and ρ are both real numbers, we have from (4)

$$u = \log \rho = \log | z |.$$

It is to be noticed that for $z = 0$, and therefore $\rho = 0$, the equation $e^u = \rho$ has no finite solution. For all other values of z, the equations

$$u = \log \rho = \log | z |,$$
$$v = \phi = \text{amp}\, z$$

determine definite values of the coördinates u, v. The corresponding value of w is defined as the logarithm of z. We have then as the formal definition of $w = \log z$

$$\boldsymbol{\log z = \log \rho + i\phi,} \tag{6}$$

which may also be written in the form

$$\log z = \log | z | + i \text{ amp}\, z. \tag{7}$$

For any particular point of the complex plane, say z_0, there are an infinite number of values of $\log z_0$ differing from each other by some multiple of $2\pi i$. This result is a consequence of the periodicity of the exponential function, which is the inverse of the logarithmic function; or, it follows directly from the definition of a logarithm, for since we have the same point z if ϕ is replaced by $(\phi + 2k\pi)$, where $k = 1, 2, \ldots$, it follows from the definition that $\log z$ has an infinite number of values for this same value of z. In

the discussions of the present chapter, unless otherwise stated, ϕ will be restricted to the chief amplitude of z, and for such a value of ϕ, $\log \rho + i\phi$ is called the **principal value** of the logarithm. In a subsequent chapter we shall discuss the logarithm as a multiple-valued function, thus giving to ϕ all possible values.

The logarithms of the positive real numbers appear as a special case of those of complex numbers, because for such numbers the value of ϕ is zero. The logarithms of negative numbers may now be given a definite significance; for, if z is a negative real number, we have

$$z = \rho(\cos \pi + i \sin \pi),$$

and hence we obtain

$$\log z = \log \rho + i\pi,$$

which is represented by a definite point in the complex plane.

Except for $z = 0$, the logarithmic function is holomorphic in the finite region; for, the Cauchy-Riemann differential equations are satisfied. We have

$$\begin{aligned} w = u + iv &= \log \rho + i\phi \\ &= \log \sqrt{x^2 + y^2} + i \operatorname{arc} \tan \frac{y}{x}. \end{aligned}$$

Consequently, we obtain

$$u = \log \sqrt{x^2 + y^2}, \qquad v = \operatorname{arc} \tan \frac{y}{x},$$

whence

$$\frac{\partial u}{\partial x} = \frac{x}{x^2 + y^2}, \qquad \frac{\partial v}{\partial x} = -\frac{y}{x^2 + y^2},$$

$$\frac{\partial u}{\partial y} = \frac{y}{x^2 + y^2}, \qquad \frac{\partial v}{\partial y} = \frac{x}{x^2 + y^2}.$$

Hence, with the restriction placed upon its amplitude the function is analytic. This conclusion does not hold, however, for $x = 0$, $y = 0$, since the partial derivatives $\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial v}{\partial x}, \frac{\partial v}{\partial y}$ become indeterminate in this case.

In putting

$$\operatorname{amp} z \equiv \phi = \operatorname{arc} \tan \frac{y}{x},$$

it should be noted that care must be taken to distinguish between $\operatorname{arc} \tan \frac{-b}{a}$ and $\operatorname{arc} \tan \frac{b}{-a}$. We may not replace these two ex-

pressions by arc $\tan\left(-\frac{b}{a}\right)$, as one might at first think possible; for, the first expression is the amplitude of $z = a - ib$, while the second is the amplitude of $z = -a + ib$. These two values of z have the same moduli, but their amplitudes differ by π. For a similar reason, we must distinguish between arc $\tan\frac{-b}{-a}$ and arc $\tan\frac{b}{a}$.

The function $\log z$ obeys the laws of logarithms for real variables. We have, for example,

$$\log z_1 + \log z_2 = \log (z_1 z_2), \tag{8}$$

where z_1, z_2 are different from zero.

To show that this relation holds, we have

$$\begin{aligned}\log z_1 + \log z_2 &= \{\log \rho_1 + i\phi_1\} + \{\log \rho_2 + i\phi_2\}\\ &= \{\log \rho_1 + \log \rho_2\} + \{i\phi_1 + i\phi_2\}\\ &= \log \rho_1\rho_2 + i(\phi_1 + \phi_2)\\ &= \log (z_1 z_2);\end{aligned}$$

since, in multiplying two complex numbers we multiply the moduli and add the amplitudes.

To find the derivative of a logarithm, we have

$$\frac{dw}{dz} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x},$$

and therefore

$$\begin{aligned}\frac{d \log z}{dz} &= \frac{\partial}{\partial x}\log\sqrt{x^2 + y^2} + i\frac{\partial}{\partial x}\text{arc}\tan\frac{y}{x}\\ &= \frac{x}{x^2 + y^2} - \frac{iy}{x^2 + y^2} = \frac{1}{x + iy} = \frac{1}{z}.\end{aligned} \tag{9}$$

The mapping of the Z-plane upon the W-plane by means of the logarithmic function gives a conformal representation because the function is analytic. The only finite singular point is the origin. As we have seen the logarithm is the inverse function of the exponential function. The general properties of the configuration obtained by mapping by means of the relation $w = \log z$ follow at once from the earlier discussion of the mapping by means of the functional relation $w = e^z$. The only difference in the two cases is that the Z-plane and the W-plane are interchanged.

Because the logarithmic function enables us to map a system of concentric circles and their orthogonal rays into a system of parallel lines and another system of straight lines orthogonal to them, it is for some purposes one of the most useful of the functions thus far discussed. One of the important applications of this function is to that system of map drawing known as **Mercator's projection.** If we undertake to map the earth's surface upon a plane by means of a projection from the north pole as the center of projection, we obtain a corresponding configuration in the plane. The region about the north pole becomes very much distorted by this process. It is often desirable in navigation so to direct a ship's course as to cut the successive meridians at a given angle, that is, to keep the ship headed toward a definite point of the compass. The curves indicating the ship's course upon the earth's surface, known as **loxodromes,** project into logarithmic spirals upon the complex plane, when that plane is tangent to the earth's surface at one of the poles while the opposite pole is taken as the center of projection. By means of the logarithmic function, the system of circles and orthogonal rays into which the earth's surface maps by this method of projection may be changed into two systems of parallel straight lines orthogonal to each other. In this new system the meridians become a system of parallel straight lines, while the parallels of latitude become a system of parallels orthogonal to the first. All loxodromic curves become straight lines cutting these two systems of parallels at a given constant angle. This result simplifies the problems of navigation by the use of the mariner's compass.

The configuration represented in Fig. 59 arises in mathematical physics whenever we have a source at the origin and a sink at an infinite distance. The rays are in this case the lines of flow and the concentric circles are lines of equal velocity-potential. If we pass through the origin a straight wire of indefinite length, through which a current of electricity is passed, we have in any plane perpendicular to this wire an induced magnetic field likewise represented by the configuration in Fig. 59, except that in this case the pencil of rays become the lines of equipotential and the system of concentric circles are lines of force. If the electric current flows

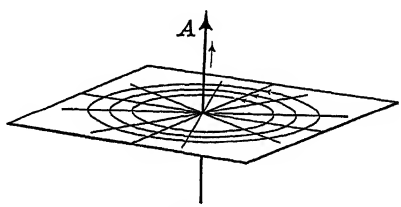

Fig. 59.

through the wire in the direction from below to above the plane of the paper, the direction of the magnetic force is then as indicated by the arrow-heads.*

Consider the function $w = \log\frac{z-1}{z+1}$. Writing the given function in the form $u + iv$, we have upon separating the real and imaginary parts

$$u = \log\frac{|z-1|}{|z+1|} = \log\frac{\sqrt{(x-1)^2+y^2}}{\sqrt{(x+1)^2+y^2}},$$

$$v = \operatorname{amp}(z-1) - \operatorname{amp}(z+1) = \arctan\frac{y}{x-1} - \arctan\frac{y}{x+1}.$$

For $u = c$, we have

$$c = \log\frac{\sqrt{(x-1)^2+y^2}}{\sqrt{(x+1)^2+y^2}},$$

$$e^{2c} = \frac{(x-1)^2+y^2}{(x+1)^2+y^2},$$

or

$$x^2 + y^2 + 2\frac{e^{2c}+1}{e^{2c}-1}x + 1 = 0. \tag{10}$$

This equation is represented by a system of coaxial circles having their centers on the X-axis. The centers of these circles are given by the equations

$$x = -\frac{e^{2c}+1}{e^{2c}-1}, \quad y = 0,$$

and the radii are equal to

$$\sqrt{\left(\frac{e^{2c}+1}{e^{2c}-1}\right)^2 - 1}.$$

For negative values of c the center lies to the right of the origin. The point $(+1, 0)$ lies within all of these circles. For positive values of c the centers of all the circles lie to the left of the origin and inclose the point $(-1, 0)$. Corresponding to $c = 0$, we have a circle of infinite radius, that is a straight line perpendicular to the segment joining the points $(1, 0)$ and $(-1, 0)$ at its middle point, namely at the origin.

For $v = c$, we have

$$c = \arctan\frac{y}{x-1} - \arctan\frac{y}{x+1},$$

whence

$$\tan c = \frac{\dfrac{y}{x-1} - \dfrac{y}{x+1}}{1 + \dfrac{y^2}{x^2-1}} = \frac{2y}{x^2+y^2-1},$$

or

$$x^2 + y^2 - \frac{2}{\tan c}y = 1. \tag{11}$$

* See J. J. Thomson, *Electricity and Magnetism*, 4th Ed., p. 329.

This is the equation of a system of coaxial circles having their centers on the Y-axis. Each of these circles passes through the points $(1, 0)$ and $(-1, 0)$. The general form of the configuration in the Z-plane is given in Fig. 60.

This configuration may be reproduced in the physical laboratory as follows. Given a glass plate covered with iron filings. Pass long straight parallel wires through the points $A \equiv (+1, 0)$ and $B \equiv (-1, 0)$ perpendicular to the plate

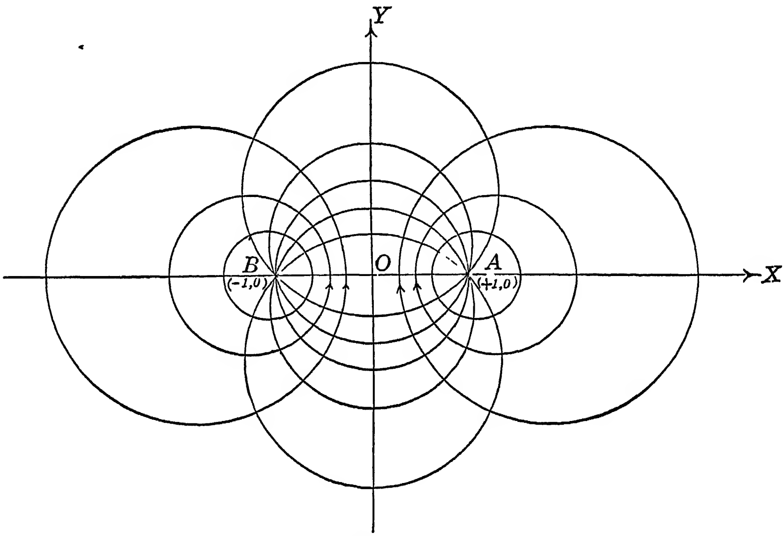

FIG. 60.

and allow an electric current of equal strength to flow in opposite directions through the two wires. By jarring slightly, the filings tend to arrange themselves along the system of circles about the points A, B and having their centers upon axis of reals. These circles are the lines of magnetic force. The direction of this force depends upon the direction of the two currents. If the direction of the current through the plane of the paper at A is downward and that through B is upward, the direction of the force is as indicated in the figure. The orthogonal system of circles all pass through the points A, B and are the lines of equipotential.

For the flow of incompressible fluids, we obtain the same configuration whenever one of the points A, B is a source and the other a sink, both being of the same strength. The circles through A and B are then the lines of flow and the circles of the orthogonal system are the lines of equal velocity-potential. If A is the source and B the sink then the direction of the flow is from A to B.

For the function $w = \log (z + 1)(z - 1)$, we have a different configuration. From the given function, we obtain

$$u + iv = \log |z + 1| \cdot |z - 1| + i\{\operatorname{amp}(z + 1) + \operatorname{amp}(z - 1)\},$$

or
$$u = \log |z+1| \cdot |z-1| = \log \sqrt{(x+1)^2+y^2}\,\sqrt{(x-1)^2+y^2},$$
$$v = \operatorname{amp}(z+1) + \operatorname{amp}(z-1) = \arctan \frac{y}{x+1} + \arctan \frac{y}{x-1}.$$

For $u = c$, we get
$$x^4 + y^4 + 2(x^2+1)\,y^2 - 2\,x^2 + 1 = e^{2c},$$

which for $c \neq 0$ is represented by a system of Cassinian ovals as shown in Fig. 61.

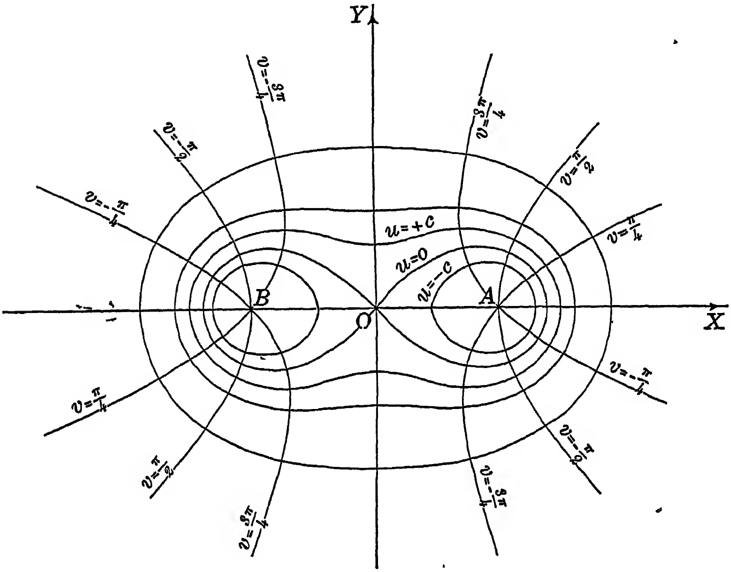

FIG. 61.

For $c = 0$ the equation represents a lemniscate having its double point at the origin. For the orthogonal system of curves, we have
$$c = \arctan \frac{y}{x+1} + \arctan \frac{y}{x-1},$$

whence
$$\tan c = \frac{\dfrac{y}{x+1} + \dfrac{y}{x-1}}{1 - \dfrac{y^2}{x^2-1}} = \frac{2\,x\,y}{x^2 - y^2 - 1},$$

or
$$x^2 - y^2 - \frac{2}{\tan c}\,xy = 1.$$

This equation gives a system of hyperbolas passing through the two points $(1, 0)$ and $(-1, 0)$ as indicated in Fig. 61.

This configuration comes into consideration in theoretical physics whenever the points A and B are sources of equal strength. If we are considering the flow

of an incompressible fluid, the curves $u = c$ are the lines of equal velocity-potential, and the curves $v = c$ are the lines of flow. If, however, we are considering a magnetic field, induced by passing currents of equal strength in the same direction through parallel straight wires intersecting the plane at A and B, the curves $u = c$ become the lines of force, and $v = c$ are the lines of equipotential. Ordinarily a line of force does not intersect itself. In the case under consideration one of the lines of force, namely $u = 0$, does intersect itself, having a double point at the origin. In order that a double point may exist the partial derivatives of u with respect to x and y must vanish.* These partial derivatives are the components of the force acting, and since both are zero there can be no force at such a point. For this reason such a point is called a **point of equilibrium.**† In the case of an irrotational fluid motion the components of the velocity are zero at a point of equilibrium and hence no flow takes place at such a point. The same configuration occurs in the discussion of the colored rings in biaxial crystals due to the interference of polarized light.

By means of the logarithmic function, we may express the more general case of any number of sources and any number of sinks, each having a given strength. Suppose we have a source at each of the points $\alpha_1, \alpha_2, \ldots, \alpha_n$ having strengths of $k_1, k_2, \ldots, k_n$, respectively. Let there be a sink at each of the points $\beta_1, \beta_2, \ldots, \beta_m$, each of strength $\lambda_1, \lambda_2, \ldots, \lambda_m$, respectively. Since the sinks are to be considered as negative sources, the corresponding factors appear in the denominator of the function of which the logarithm is to be taken. The corresponding function is then

$$w = \log \frac{(z - \alpha_1)^{k_1}(z - \alpha_2)^{k_2} \ldots (z - \alpha_n)^{k_n}}{(z - \beta_1)^{\lambda_1}(z - \beta_2)^{\lambda_2} \ldots (z - \beta_m)^{\lambda_m}}.$$

As a special case which presents some interest, let us consider the function

$$w = \log \frac{(z - 1)^2}{z + 1}.$$

The function w determines, for example, the equipotential lines and the lines of force in a magnetic field about two parallel conductors in which the electric current is passing in opposite directions in the two and is twice as strong in the one as in the other. The wires pierce the complex plane at $A \equiv (1, 0)$ and $B \equiv (-1, 0)$. The wire through A carries a current twice as strong as the one through B.

In order to obtain the two systems of conjugate curves given by the function, put

$$z - 1 = \rho_1 e^{i\theta_1}, \qquad z + 1 = \rho_2 e^{i\theta_2};$$

* See Townsend and Goodenough, *First Course in Calculus*, p. 370.

† See Maxwell, *Electricity*, Vol. I, Chap. VI; Jeans, *Electricity and Magnetism*, p. 59; Lamb, *Hydrodynamics*, p. 17.

we have

$$w = \log \frac{\rho_1^2 e^{2i\theta_1}}{\rho_2 e^{i\theta_2}} = \log \frac{\rho_1^2}{\rho_2} + \log e^{i(2\theta_1 - \theta_2)}$$

$$= \log \frac{\rho_1^2}{\rho_2} + i(2\theta_1 - \theta_2).$$

Hence, we obtain

$$u = \log \frac{\rho_1^2}{\rho_2}, \qquad v = 2\theta_1 - \theta_2.$$

For $u = c$, we have

$$c = \log \frac{\rho_1^2}{\rho_2},$$

or

$$\rho_2 = \frac{\rho_1^2}{e^c} = K\rho_1^2, \qquad K > 0. \tag{12}$$

For the orthogonal system, we have

$$c = 2\theta_1 - \theta_2. \tag{13}$$

To plot any one of the system of curves represented by (12), give K an assigned value and give to ρ_1 any convenient succession of values. Compute the corresponding values of ρ_2 by means of (12). With $z = +1$ as a center and the assumed values of ρ_1 as radii, draw circles. Likewise, with $z = -1$ as a center and the computed values of ρ_2 as radii draw circles. The intersections of corresponding circles give points on the required curve.

To plot a curve belonging to the system given by (13), give to c any assigned value and from the points $z = +1$, $z = -1$, and draw lines making angles θ_1 and $\theta_2 = 2\theta_1 - c$, respectively, with the axis of reals. The intersection of corresponding lines gives points on the required curve. The general form of the two systems of curves is shown in Fig. 62.

To determine the double points of the lines of force, that is, the points of equilibrium, we have

$$u = \log \frac{\rho_1^2}{\rho_2} = \log \frac{(x-1)^2 + y^2}{\sqrt{(x+1)^2 + y^2}}.$$

The double points are given by putting partial derivatives of u with respect to x and y equal to zero and solving the two resulting equations for x and y. We have then to solve the equations

$$\frac{\partial u}{\partial x} = \frac{\partial}{\partial x} \log \frac{(x-1)^2 + y^2}{\sqrt{(x+1)^2 + y^2}} = \frac{2(x-1)}{(x-1)^2 + y^2} - \frac{x+1}{(x+1)^2 + y^2} = 0, \tag{14}$$

$$\frac{\partial u}{\partial y} = \frac{\partial}{\partial y} \log \frac{(x-1)^2 + y^2}{\sqrt{(x+1)^2 + y^2}} = \frac{2y}{(x-1)^2 + y^2} - \frac{y}{(x+1)^2 + y^2} = 0. \tag{15}$$

Equations (14) and (15) are satisfied simultaneously by the values $y = 0$, $x = -3$. These values are therefore the coördinates of the point C of equilibrium. To

determine which one of the lines $u = c$ maps into the particular curve having a double point at $(-3, 0)$, we substitute the values $x = -3$, $y = 0$ in (12) and determine the corresponding value of c. This substitution gives

$$e^{2c} = 64, \qquad \text{or} \qquad c = \log 8;$$

that is, the potential function has at each point of this curve the value log 8.

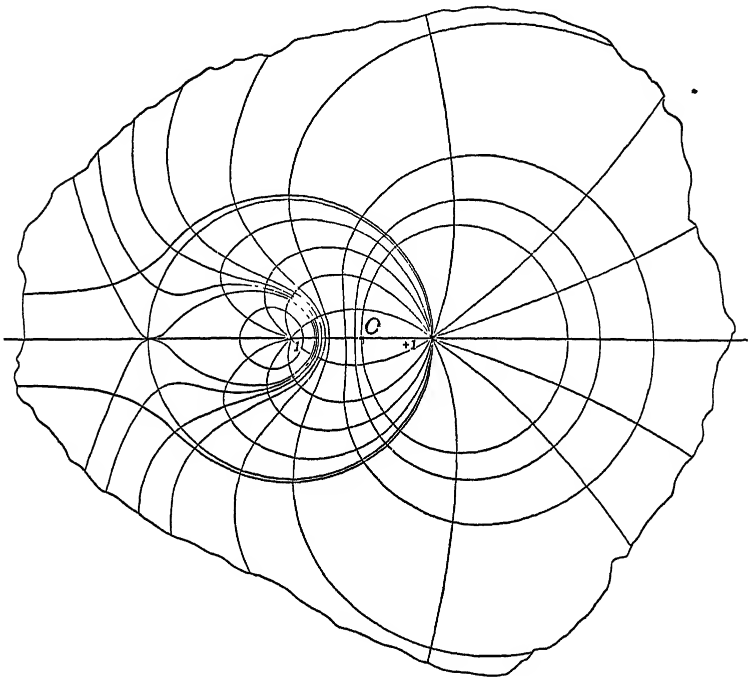

FIG. 62.

The distribution of matter being confined to the plane, the intensity of the force at any point per unit of strength is equal to the reciprocal of the distance.* Hence, in order that C shall be a point of equilibrium, it follows from the laws of physics that C must lie on the X-axis and that we must have

$$\frac{2}{AC} - \frac{1}{BC} = 0.$$

It will be seen that this equation gives the same values of the coördinates of C as those already obtained.

* See Wangerin, *Theorie des Potentials und der Kugelfunktionen,* Vol. I, pp. 135–137.

31. Trigonometric Functions. The definition of the various trigonometric functions may be made to depend upon the exponential function e^z already defined. From the definition of e^z, it was shown that

$$e^{i\theta} = \cos\theta + i\sin\theta,$$

whence

$$e^{-i\theta} = \cos\theta - i\sin\theta,$$

where θ in both cases is real. Solving these equations for $\sin\theta$, $\cos\theta$, we get

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i},$$

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}.$$

In a similar manner, we shall now define $\sin z$, $\cos z$ in terms of the exponential function e^z, by putting

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i},$$

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}.$$

Since the function e^z is analytic, it follows that $\sin z$, $\cos z$ are also analytic functions. Moreover, $\sin x$ and $\cos x$ appear as special cases of the sine and cosine of the complex variable z.

The trigonometric functions of a complex variable satisfy the same trigonometric identities as the corresponding functions of real variables. We may show, for example, that the following relation holds

$$\sin(z_1 + z_2) = \sin z_1 \cos z_2 + \cos z_1 \sin z_2.$$

We have

$$\begin{aligned}
\sin z_1 \cos z_2 + \cos z_1 \sin z_2 &= \frac{(e^{iz_1} - e^{-iz_1})(e^{iz_2} + e^{-iz_2})}{2i \cdot 2} \\
+ \frac{(e^{iz_1} + e^{-iz_1})(e^{iz_2} - e^{-iz_2})}{2 \cdot 2i} &= \frac{2e^{i(z_1+z_2)} - 2e^{-i(z_1+z_2)}}{4i} \\
&= \frac{e^{i(z_1+z_2)} - e^{-i(z_1+z_2)}}{2i} \\
&= \sin(z_1 + z_2).
\end{aligned}$$

The remaining trigonometric identities may be established in a similar manner. While the fundamental identity

$$\cos^2 z + \sin^2 z = 1$$

holds for complex as well as real values of z, it follows from the definitions of $\sin z$, $\cos z$, that by the proper choice of the complex variable z either of the functions $\sin z$ and $\cos z$ can be made greater than unity in absolute values, thus differing in this respect from the case where the variable is real.

Since e^{iz} has the period 2π, it follows from the definition of $\sin z$ that it likewise has the period 2π; for, we have

$$\begin{aligned}\sin(z+2\pi) &= \frac{e^{i(z+2\pi)} - e^{-i(z+2\pi)}}{2i} \\ &= \frac{e^{iz} - e^{-iz}}{2i} \\ &= \sin z.\end{aligned}$$

In a similar manner it may be shown that $\cos z$ has the period 2π; that is, that

$$\cos(z+2\pi) = \cos z.$$

The remaining trigonometric functions are periodic, having the same periodicity as the corresponding functions of a real variable. As we shall see the lines that limit the fundamental region of $\sin z$ and $\cos z$ are parallel to the Y-axis, while the fundamental region for e^z is bounded by lines parallel to the axis of reals. This difference is a consequence of inserting the factor i before z in the definition of $\sin z$, $\cos z$. It is to be noted also that for the exponential functions e^z, e^{iz} the region of periodicity is identical with the fundamental region; that is to say, no two points z in one of the strips defining the region of periodicity gives the same value of the function. In the case of $\sin z$ and $\cos z$ the situation is different and each of these functions has the same value for two different values of z in the strip defining the region of periodicity. For example, if we substitute $\pi - z$ for z in

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i},$$

we have

$$\begin{aligned}\sin(\pi - z) &= \frac{e^{i(\pi-z)} - e^{-i(\pi-z)}}{2i} \\ &= \frac{e^{i\pi}e^{-iz} - e^{-i\pi}e^{iz}}{2i}.\end{aligned}$$

Remembering that

$$\begin{aligned}e^{i\pi} &= \cos\pi + i\sin\pi = -1, \\ e^{-i\pi} &= \cos\pi - i\sin\pi = -1,\end{aligned}$$

we have
$$\sin(\pi - z) = \frac{-e^{-iz} + e^{iz}}{2i}$$
$$= \sin z.$$

Moreover, the points representing z and $\pi - z$ both lie within the region of periodicity $-\pi < x \leqq \pi$, as shown in Fig. 63, provided the real part of z is greater than zero and not greater than π. This result

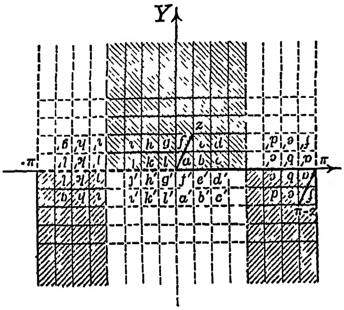

FIG. 63.

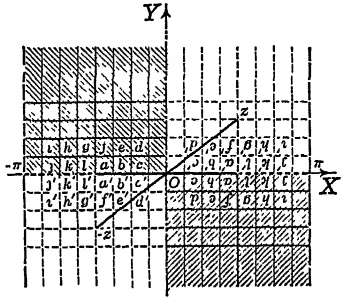

FIG. 64.

shows that the region of periodicity of $\sin z$ can not be taken as a fundamental region of that function.

Again, if in
$$\cos z = \frac{e^{iz} + e^{-iz}}{2}$$
we replace $-z$ by z, we have the same function as before. Hence, we may write
$$\cos(-z) = \cos z.$$

If z is represented by a point in the upper right-hand portion of the strip of periodicity, as shown in Fig. 64, then $-z$ is a point in the lower left-hand portion as shown. The two points lie symmetrically with respect to the origin. This fact shows that the region of periodicity $-\pi < x \leqq \pi$ does not answer the purpose of a fundamental region for $\cos z$.

Neither the sine nor the cosine can have the same value for more than two values of z in the same periodic strip. For example, for the cosine, we have
$$\cos z = \frac{e^{iz} + e^{-iz}}{2}$$
$$= \frac{e^{2iz} + 1}{2e^{iz}}, \tag{1}$$

which is of the second degree in e^{iz}; and hence if we put $\cos z$ equal to a constant there are but two values of e^{iz} and hence of z that satisfy this equation.

As we have already seen, the regions of periodicity can not always be taken as the fundamental region. We shall show that the region bounded by the lines $x = 0$, $x = \pi$ may be taken as the fundamental region for $\cos z$. In this connection, we shall consider the mapping of the Z-plane upon the W-plane by means of the relation $w = \cos z$. We have

$$\cos z = w = u + iv = \frac{e^{iz} + e^{-iz}}{2}$$
$$= \frac{e^{i(x+iy)} + e^{-i(x+iy)}}{2}$$
$$= \frac{e^{ix}e^{-y} + e^{-ix}e^{y}}{2}$$
$$= \frac{e^{-y}}{2}(\cos x + i \sin x) + \frac{e^{y}}{2}(\cos x - i \sin x)$$
$$= \frac{e^{-y} + e^{y}}{2}\cos x + i\frac{e^{-y} - e^{y}}{2}\sin x.$$

Hence, we may write

$$u = \frac{e^{-y} + e^{y}}{2}\cos x, \qquad v = \frac{e^{-y} - e^{y}}{2}\sin x. \tag{2}$$

For $x = 0$, we obtain

$$u = \frac{e^{-y} + e^{y}}{2}, \qquad v = 0.$$

If $y = 0$, we have u equal to $+1$; for $y \lessgtr 0$, we get $u > 1$. Hence, the axis of imaginaries maps into that portion of the U-axis that lies to the right of $u = 1$, as shown in Fig. 65.

For $x = \pi$, we have

$$u = -\frac{e^{-y} + e^{y}}{2}, \qquad v = 0,$$

and the line $x = \pi$ maps into that portion of the negative U-axis that lies to the left of the point -1. If $0 < x < \pi$, we have for a positive value of y a corresponding negative value of v; for, the value of $\sin x$ is positive, while the factor

$$\frac{e^{-y} - e^{y}}{2}$$

is negative. In a similar way, if y is negative and $0 < x < \pi$, the corresponding point in the W-plane lies above the axis of reals.

It will be seen upon inspection that for $x = \frac{\pi}{2}$, $y = 0$, the value of w is zero. As x varies from 0 to π, y remaining zero, w takes all of

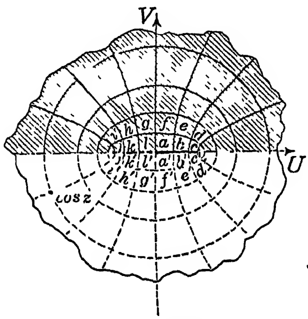

FIG. 65.

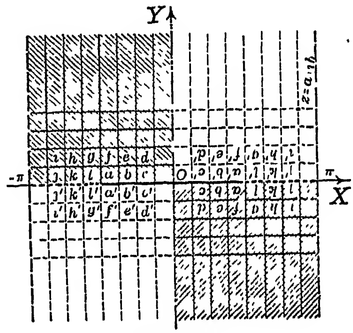

FIG. 66.

the values represented by points on the real axis between $+1$ and -1; for, in this case we have from (2)

$$u = \cos x, \qquad v = 0.$$

Lines parallel to the X-axis map into ellipses (Fig. 65) having the points ± 1 as the common foci. We may show this as follows. From (2) we get

$$\cos x = \frac{2u}{e^y + e^{-y}}, \qquad \sin x = \frac{-2v}{e^y - e^{-y}}.$$

Squaring both members of these equations and adding, we have

$$\left(\frac{2u}{e^y + e^{-y}}\right)^2 + \left(\frac{2v}{e^y - e^{-y}}\right)^2 = 1.$$

For $y = c$ this equation is that of an ellipse. For various values of c we obtain a system of ellipses having the common foci $+1$, -1.

The lines parallel to the Y-axis, that is $x = c$, map into hyperbolas having the foci ± 1. To get the equations of these hyperbolas, we divide the members of the first equation in (2) by $\cos x$ and those of the second by $\sin x$, thus obtaining

$$\frac{u}{\cos x} = \frac{e^{-y} + e^y}{2}, \qquad \frac{v}{\sin x} = \frac{e^{-y} - e^y}{2}.$$

Squaring these results, we have

$$\left(\frac{u}{\cos x}\right)^2 = \frac{e^{-2y} + 2 + e^{2y}}{4}, \qquad \left(\frac{v}{\sin x}\right)^2 = \frac{e^{-2y} - 2 + e^{2y}}{4}.$$

Subtracting the second of these equations from the first, we obtain

$$\left(\frac{u}{\cos x}\right)^2 - \left(\frac{v}{\sin x}\right)^2 = 1,$$

which for constant values of x is the equation required.

The region bounded by the lines $x = 0$, $x = \pi$ maps into the entire W-plane and may therefore be taken as the fundamental region for $\cos z$. Any region bounded by the lines $x = k\pi$, $x = (k + 1)\pi$, $k = \cdots, -3, -2, -1, 0, +1, +2, +3, \ldots$ answers equally well as a fundamental region.

The corresponding regions in the two planes are indicated by the letters $a, b, c \ldots$ and $a', b', c', \ldots$, Figs. 65 and 66.

The configuration in the W-plane (Fig. 65) gives us a method of determining $\cos z$ by graphical methods. For example, let $z \equiv a + ib$ be any point in the Z-plane. Suppose the parallels to the axes map into the particular ellipse and hyperbola shown; then $\cos z$ is represented by the intersection of the curves as indicated.

From the definitions of $\sin z$ and $\cos z$, we can readily obtain expressions for the other trigonometric functions in terms of the exponential function.

For example, we have

$$\tan z = \frac{\sin z}{\cos z}, \qquad \cot z = \frac{\cos z}{\sin z}, \text{ etc.}$$

These functions are holomorphic in that portion of the finite plane for which they are defined. The first of these functions, $\tan z$, is undefined for those values of z for which $\cos z$ vanishes. From the map of $\cos z$ upon the W-plane, it will be seen that $\cos z$ is equal to zero only for the real values

$$z = \frac{\pi}{2} \pm k\pi, \qquad k = 0, 1, 2, \ldots.$$

In a similar manner, it may be shown that $\cot z$ is undefined for those values of z for which $\sin z$ vanishes, that is for the real values

$$z = \pm k\pi, \qquad k = 0, 1, 2, \ldots.$$

The trigonometric functions are therefore analytic functions.

32. Hyperbolic Functions. As in the case of circular functions, we shall first define the functions

$$w = \sinh z, \qquad w = \cosh z,$$

and from these definitions deduce the remaining functions by means of the relations,

$$\tanh z = \frac{\sinh z}{\cosh z}, \qquad \coth z = \frac{1}{\tanh z}, \qquad \operatorname{sech} z = \frac{1}{\cosh z},$$
$$\operatorname{cosech} z = \frac{1}{\sinh z}.$$

We now define $\sinh z$, $\cosh z$ in terms of the exponential function as follows:

$$\sinh z = \frac{e^z - e^{-z}}{2}, \qquad \cosh z = \frac{e^z + e^{-z}}{2}. \tag{1}$$

By comparing these definitions with those of the sine and cosine, it will be seen at once that

$$\sinh z = -i \sin iz, \qquad \cosh z = \cos iz. \tag{2}$$

The following useful identities follow at once from the definitions given.

$$\cosh^2 z - \sinh^2 z = 1,$$
$$\operatorname{sech}^2 z + \tanh^2 z = 1,$$
$$\coth^2 z - \operatorname{cosech}^2 z = 1.$$

To deduce the first relation, we have

$$\begin{aligned}\cosh^2 z - \sinh^2 z &= \left(\frac{e^z + e^{-z}}{2}\right)^2 - \left(\frac{e^z - e^{-z}}{2}\right)^2 \\ &= \frac{e^{2z} + 2 + e^{-2z}}{4} - \frac{e^{2z} - 2 + e^{-2z}}{4} \\ &= 1.\end{aligned}$$

The rest of the above identities may be deduced in a similar manner.

The hyperbolic functions are analytic functions, since e^z is an analytic function. Moreover these functions are periodic; for, as we have seen the function e^z is periodic.

The formulas for hyperbolic functions of real variables may be

extended without change to complex variables. We have, for example,

$$\begin{aligned}\cosh(z_1+z_2) &= \cos i(z_1+z_2)\\ &= \cos iz_1 \cos iz_2 - \sin iz_1 \sin iz_2\\ &= \cosh z_1 \cosh z_2 + \sinh z_1 \sinh z_2.\end{aligned}$$

Hyperbolic functions of real variables may often be conveniently used to express the trigonometric functions of a complex variable in the form

$$u(x, y) + iv(x, y).$$

We have, for example,

$$\begin{aligned}\sin z = \sin(x+iy) &= \sin x \cdot \cos iy + \cos x \cdot \sin iy\\ &= \sin x \cdot \cosh y + i \cos x \cdot \sinh y,\end{aligned}$$

whence $$u = \sin x \cosh y, \quad v = \cos x \sinh y.$$

Similarly $$\cos z = \cos x \cosh y - i \sin x \sinh y,$$

and $$u = \cos x \cosh y, \quad v = -\sin x \sinh y.$$

To express $\tan z$ in the form $u + iv$, we have

$$\begin{aligned}\tan z = \frac{\sin z}{\cos z} &= \frac{\sin(x+iy)}{\cos(x+iy)}\\ &= \frac{\sin(x+iy)\cos(x-iy)}{\cos(x+iy)\cos(x-iy)}\\ &= \frac{\sin 2x + \sin 2iy}{\cos 2x + \cos 2iy} = \frac{\sin 2x + i \sinh 2y}{\cos 2x + \cosh 2y},\end{aligned}$$

whence $$u = \frac{\sin 2x}{\cos 2x + \cosh 2y}, \qquad v = \frac{\sinh 2y}{\cos 2x + \cosh 2y}.$$

We shall now map the Z-plane upon the W-plane by means of the relation

$$w = \cosh z.$$

The results of this mapping can be readily deduced from those obtained in mapping by means of the relation $w = \cos z$; for, it will be seen from (2) that we have $w = \cosh z$ if in $w = \cos z$, z is replaced by iz. Hence, to map any configuration from the Z-plane to the W-plane by means of the relation $w = \cosh z$, all that is necessary is first to map the given configuration from the Z-plane to an auxiliary Z'-plane by means of the relation $z' = iz$, which merely rotates each point of the complex plane through a positive angle $\frac{\pi}{2}$, and then to map the resulting configuration from

the Z'-plane to the W-plane by means of the relation $w = \cos z'$. The region in the Z'-plane bounded by the lines $x' = 0$, $x' = \pi$ may be regarded as the fundamental region for the function $w = \cos z'$. This region corresponds to the region in the Z-plane bounded by the lines $y = 0$, $y = -\pi$, which may therefore be taken as the fundamental region for the function $w = \cosh z$. As may be seen, any one of the regions bounded by the lines

$$y = k\pi, \quad y = (k-1)\pi, \quad k = \cdots, -2, -1, 0, 1, 2, \ldots$$

can be used as a fundamental region.

A system of lines in the Z-plane parallel to either of the coördinate axes maps by means of the relations

$$w = \cosh z, \quad w = \cos z$$

into a system of straight lines in the W-plane which are likewise parallel to the coördinate axes. The lines that map in the one case into ellipses map in the other case into hyperbolas and conversely. This result is verified by a comparison of the equations for u and v in the two cases. We obtain from $w = u + iv = \cosh z$

$$u = \cosh x \cos y = \frac{e^x + e^{-x}}{2} \cos y,$$

$$v = \sinh x \sin y = \frac{e^x - e^{-x}}{2} \sin y,$$

which are the same equations as those obtained from $w = \cos z$, Art. 31, except that x is replaced by y and y by $-x$. In other words by the change from $\cos z$ to $\cosh z$ the lines of level and the lines of slope are interchanged.

In a similar manner we may establish relations between the maps obtained by means of the remaining circular functions and the corresponding hyperbolic functions. From (2) it will be seen that the introduction of the factor i enables us to express any hyperbolic function in terms of the corresponding circular function. Consequently, it follows that the special significance of hyperbolic functions is confined to functions of a real variable.

Because of the similarity of the configurations obtained by mapping the lines $x = c$, $y = c$, by means of the relations $w = \cos z$ and $w = \cosh z$, it is to be expected that similar applications may be made in theoretical physics. If we have the case of a liquid flowing about an elliptic cylinder whose intersection by the complex plane is an ellipse having its foci at -1 and $+1$, respectively, then the ellipses, Fig. 65, are the lines of flow and the hyperbolas are the lines of equal

velocity-potential. As a limiting case we have the flow of a liquid about a thin plate joining the points $+1$ and -1.* If that portion of the positive real axis lying to the right of $+1$ be regarded as a line source and that portion of the negative real axis lying to the left of -1 be taken as a sink the ellipses are again the lines of flow and the hyperbolas equipotential lines. If, however, the line joining $+1$ and -1 is regarded as a source, then the hyperbolas are the lines of flow and the ellipses are the equipotential lines.

The definitions of the transcendental functions thus far discussed have been based upon the definition of e^z, which in turn was defined in terms of known functions of real variables. Other methods of procedure could have been employed. For example, the logarithm of z could have been, and often is, defined as the integral $\int_1^z \frac{dz}{z}$. From this definition the properties of a logarithm can be readily developed. Then e^z may be defined as the inverse function of $\log z$, and the remaining functions can be defined as in the text. The other transcendental function which we have given may also be defined by means of integrals; for example, we may make use of the following relations as definitions

$$\text{arc}\tan z = \int_0^z \frac{dz}{1+z^2}, \qquad \text{arc}\sin z = \int_0^z \frac{dz}{\sqrt{1-z^2}},$$

upon which the definitions of the remaining functions discussed in the text may be based.

EXERCISES

1. Discuss the conjugate functions determined by the relation

$$w^2 = z + 1.$$

Plot the projections upon the XY-plane of the lines of level and lines of slope.

2. Discuss the mapping upon the W-plane of a system of concentric circles about the origin in the Z-plane, by means of the relation

$$w = 3z^3 + 5.$$

3. Show that the function

$$z = r\{\cos \lambda t + i \sin \lambda t\},$$

where t is the independent variable representing time and where r and λ are real constants, represents a movement of the z-point such that the velocity ν of the z-point is constant in magnitude but varying in direction, and such that the acceleration of the z-point is always directed toward the origin and is constant in magnitude and equal to $\frac{\sigma^2}{r}$, σ being the absolute value of ν.

* See Lamb, *Hydrodynamics*, 3d Ed., p. 69; Webster, *Electricity and Magnetism*, p. 319.

4. Given $w = e^z$. Suppose that the z-point has a constant real velocity σ along the line $y = \phi$ in the Z-plane. Show, by differentiation, that the corresponding point of the W-plane moves along the straight line $u = v \cdot \cot \phi$ with a varying speed.

5. Any straight line through the origin making an angle different from zero with the X-axis crosses an infinite number of fundamental regions of the function $w = e^z$. Explain the fact that such a line maps into a single continuous curve in the W-plane.

6. Given $w = \left(1 + \frac{z}{n}\right)^n$, $n = 2, 3, \ldots$. Determine fundamental regions for this function for the various values of n and show how we may obtain a fundamental region for $w = e^z$ as the limiting case.

7. Show that e^z is an automorphic function.

8. Construct the map of the function $w = \sin z$ similar to the map of $\cos z$ shown in Figs. 65, 66.

9. Making use of Figs. 65, 66, and the figures obtained for the function $w = \sin z$, construct the corresponding figures for the functions $w = \cosh z$, $w = \sinh z$.

10. Show that $D_z \sin z = \cos z$, $D_z \sinh z = \cosh z$.

11. Show that $\sin 2z = 2 \sin z \cos z$, $\sinh 2z = 2 \sinh z \cosh z$.

12. Show that for $w = \sinh z$, we have $u = \sinh x \cos y$, $v = \cosh x \sin y$.

13. Show that for $w = \cosh z$, we have $u = \cosh x \cos y$, $v = \sinh x \sin y$.

14. Prove that

$$\tanh z = \frac{\sinh x \cosh x + i \sin y \cos y}{\cos^2 y \cosh^2 x + \sin^2 y \sinh^2 x}.$$

15. Discuss the mapping of orthogonal systems of straight lines parallel to the axes in the W-plane upon the Z-plane by means of the relation

$$w = \log (z - 1)(z + 1)(z - i).$$

Discuss the possible applications in theoretical physics.

16. Discuss the function

$$w = \log \frac{(z - 1)^3}{z^2},$$

and point out possible applications as suggested by the map in the Z-plane of the lines $u = c$, $v = c$. Locate the points of equilibrium, if such exist.

17. Suppose a system of equipotential curves to be given by the confocal ellipses

$$\frac{x^2}{a^2 + \lambda} + \frac{y^2}{b^2 + \lambda} = 1, \quad -a^2 < -b^2 \leqq \lambda.$$

Show that the lines of flow are the confocal hyperbolas

$$\frac{x^2}{a^2 + \lambda} + \frac{y^2}{b^2 + \lambda} = 1, \quad -a^2 < \lambda \leqq -b^2.$$

18. Show, by the method of Ex. 3, that for uniform motion along any curve, the acceleration is always directed toward the center of curvature and in magnitude is equal to

$$\frac{\sigma^2}{\text{radius of curvature}},$$

where σ is the speed in the path.

19. Show that for non-uniform motion in any curve, the component of the acceleration normal to the path is

$$\frac{\sigma^2}{\text{radius of curvature}},$$

where σ is the varying speed in the path.

20. The logarithmic function and the hyperbolic functions have been defined in terms of e^z. By means of these definitions, show that

$$z = \log\frac{1+w}{1-w} = 2 \operatorname{arc\,tanh} w.$$

21. By aid of the Cauchy integral formula, compute the value of $\sin z$ at the point $\rho = \frac{1}{2}$, $\theta = \frac{\pi}{2}$.

CHAPTER V

LINEAR FRACTIONAL TRANSFORMATIONS

33. Definition of linear fractional transformation. In several of the illustrative examples of mapping thus far considered, the relation between w and z is such that a portion of the one plane maps into the whole of the other plane. For example, in the case of $w = z^2$ one-half of the Z-plane maps into the whole of the W-plane. To each point in the W-plane there correspond then two points in the Z-plane, symmetrically situated with respect to the origin. We shall now consider the general linear algebraic relation between w and z. This relation may be written in the form

$$w = \frac{\alpha z + \beta}{\gamma z + \delta}, \tag{1}$$

where α, β, γ, δ are constants, real or complex, and the determinant

$$\begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix}$$

is different from zero. The relation (1) between w and z differs from those mentioned above in this respect that to each value of z there is, with a convention as to the point at infinity to be noted in the following article, one and only one value of w and *vice versa*.

In our discussion of mapping (Chapter IV), we examined the relation between w and z by allowing one of these variables to describe a given configuration and determining the corresponding configuration described by the other variable. The one configuration was then said to be mapped upon the other configuration by means of the given functional relation. In a similar manner a given region was frequently mapped upon another region. It was convenient to represent the w-points in one complex plane and the z-points in another, employing for this purpose two distinct sets of coördinate axes, one in each plane. In the present chapter we shall study the particular functional relation given in (1) from a somewhat different point of view. We shall represent both the z-points and the w-points in the same plane and refer them to the same set of

coördinate axes. We shall attempt to find a path along which any particular z-point may be regarded as moving in passing into the corresponding w-point. To distinguish this process from that of mapping already discussed, we shall speak of it as a transformation of the complex plane. Since equation (1) expresses w as a linear fractional function of z, we shall call the transformation to be discussed a **linear fractional transformation** of the complex plane. When the relation given in (1) reduces to the form $w = \alpha z + \beta$, we shall speak of it as a **linear transformation.** If we think merely of the results of such a transformation, we may say that a given configuration is mapped upon the resulting configuration by means of a linear fractional transformation.

34. Point at infinity. The general linear fractional relation given in (1) of the last article fails to establish a complete one-to-one correspondence between the finite points of the complex plane. For example, there is no finite value of w corresponding to the value $z = -\dfrac{\delta}{\gamma}$. To make the one-to-one correspondence complete, it is customary to assign an ideal point to the complex plane, called the **point at infinity.** To this artificial point may be associated the artificial number ∞. The complex plane is to be regarded then as closed at infinity; that is, the plane is to be considered as having but a single point at infinity.

We set up a correspondence between the point $z = -\dfrac{\delta}{\gamma}$ and the ideal point $w = \infty$, and likewise between the ideal point $z = \infty$ and point $w = \dfrac{\alpha}{\gamma}$. The one-to-one correspondence between the points of the whole complex plane by means of the general linear fractional relation becomes complete by the aid of this convention; that is, to each point that may be assigned to z, there is one and only one point that represents the corresponding value of w, and conversely.

This convention concerning the point at infinity is very convenient in other connections. When we speak of the existence of the limit $\underset{z \doteq \alpha}{L} f(z)$, where α is a finite number, it is implied that the function $f(z)$ has the same limiting value as z approaches α through all possible sets of values, that is, a limiting value that is independent of any changes in the amplitude of $z - \alpha$ as the modulus of $z - \alpha$ decreases. Similarly, a function $f(z)$ may have a unique limiting value as z becomes infinite through all possible sets of values; that is, it

may have a limiting value that is independent of any changes in the amplitude of z as the modulus of z increases beyond all finite bounds. It is convenient to denote this limit by $\underset{z=\infty}{L} f(z)$ and to speak of $z = \infty$ as the limiting point in this case.

Let us now consider the limiting value of the function

$$w = \frac{\alpha z + \beta}{\gamma z + \delta},$$

as z becomes infinite and likewise that of the inverse function

$$z = \frac{-\delta w + \beta}{\gamma w - \alpha},$$

as w becomes infinite. We have

$$\underset{z=\infty}{L}\, w = \underset{z=\infty}{L} \frac{\alpha z + \beta}{\gamma z + \delta} = \frac{\alpha}{\gamma},$$

and

$$\underset{w=\infty}{L}\, z = \underset{w=\infty}{L} \frac{-\delta w + \beta}{\gamma w - \alpha} = -\frac{\delta}{\gamma}.$$

These two results fully justify, in so far as the general linear fractional function is concerned, the convention introduced concerning the nature of the complex plane at infinity.

If we think of the values of z as represented by the points of one complex plane, called the Z-plane, and the values of w as represented by the points of another complex plane, called the W-plane, then of course an ideal point at infinity must be associated with each plane.

We may speak of the neighborhood of the point at infinity just as we speak of the neighborhood of any finite point. By such a neighborhood is understood the set of points exterior to any closed curve, for example the set of points exterior to a large circle about the origin as a center. We say also that a given region contains the point at infinity if it consists of all the points exterior to a given closed curve. In this connection the substitution $z = \frac{1}{z'}$ is of special importance. By this transformation every finite point except the origin goes over into some finite point of the plane. The neighborhood of the origin corresponds to the neighborhood of the point at infinity, and the origin itself may be said to correspond to the point at infinity.

This relation between the point at infinity and the origin affords a convenient method of investigating the properties of a function

$f(z)$ for values of z in the neighborhood of the point $z = \infty$. To do so we replace in $f(z)$ the independent variable z by $\frac{1}{z'}$ and discuss the properties of the transformed function $\phi(z')$ in the neighborhood of the point $z' = 0$. If the limit $\underset{z' \doteq 0}{L}\ \phi(z')$ exists and is equal to A, then we say that $f(z)$ has the value A at the point at infinity, and we write

$$A = \underset{z=\infty}{L}\ f(z) = f(\infty).$$

If $\phi(z')$ is continuous for $z' = 0$, we say that $f(z)$ is continuous at infinity. If $z' = 0$ is a regular point of $\phi(z')$, we say that $z = \infty$ is a regular point of $f(z)$. As z becomes infinite the function $f(z)$ may also become infinite and in such a manner that $\frac{1}{f(z)}$ approaches the limiting value zero. We say then that

$$f(\infty) = \infty.$$

35. The transformation $w = z + \beta$. Before discussing the general transformation (1) of Art. 33, we shall consider some special cases that are of particular importance and first of all let us consider the transformation

$$w = z + \beta.$$

To obtain this function from the general case, put $\alpha = \delta = 1$ and $\gamma = 0$. This relation indicates that to each number z there is added another number β. From the geometric interpretation of addition it will be seen that the z-points are transformed into the corresponding w-points by moving each z-point in the direction of the line joining the point β with the origin and to a distance equal to $|\beta|$. In what follows, it frequently will be convenient to describe a transformation of the complex plane as a continuous motion, meaning thereby that all of the points of the plane are considered as moving continuously from their initial to their final positions along a system of curves. The motion just described is called a **translation.** It will be seen that a translation of the complex plane leaves the form and size of any configuration unchanged; that is, it transforms any curve into a congruent curve.

36. The transformation $w = \alpha z$. As already stated α may be a real number or a complex number of the form

$$\alpha = \rho\,(\cos\theta + i\sin\theta).$$

We shall understand more easily the full significance of this transformation by considering first some special cases. Let us suppose for example that $\theta = 0$; that is, let α be considered as a real positive number. The result is simply a multiplication of the modulus of z by the number $\rho = |\alpha|$. Geometrically, this special form of the general transformation may be regarded as moving every point along the line passing through the given point and the origin, that is, along the half-ray from the origin on which it lies. The point moves out or in along this half-ray according as ρ is greater than or less than unity. This change affects every point of the complex plane, and we may regard the transformation as representing a motion of the points of the plane. We shall refer to such a motion as an **expansion** or **stretching.** We are concerned here with the path by which the variable point may be regarded as passing from its initial to its final position, rather than the velocity with which this motion takes place. The number ρ is called the **modulus of expansion.**

The significance of ρ may also be seen from a consideration of the derivative. As we have seen, $|D_z w|$ gives the ratio of magnification that takes place in infinitesimal elements as z varies. We have

$$|D_z w| = |D_z(\alpha z)| = |\alpha| = \rho;$$

that is, any configuration in the complex plane is magnified in this ratio. Suppose that we have a system of concentric circles about the origin and a pencil of rays passing through the origin. Each half-ray remains unchanged as a whole, although any particular point upon it is moved out or in according as the ratio of expansion is greater or less than unity. By this transformation, any portion of the plane inclosed by two half-rays and two concentric circles is transformed into another portion bounded by the same two half-rays and the concentric circles into which the first two are transformed; for example, the region (2), Fig. 67, goes over into (3). Each dimension of the region has been multiplied by the number ρ.

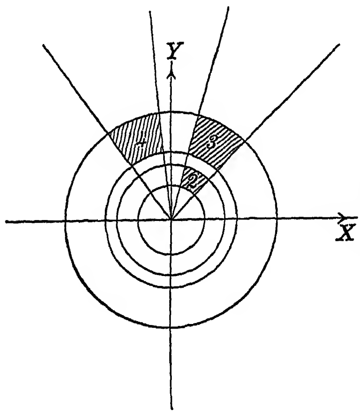

FIG. 67.

Suppose we now allow θ to vary, while ρ remains constantly equal to one. Then by the laws of multiplication already established, we

obtain from any value z the corresponding value of w by adding to the amplitude of z the angle θ. Inasmuch as $\rho = 1$, no magnification takes place and the resulting configuration is obtained by revolving each point z about the origin counter-clockwise through the angle θ. Considered from the standpoint of the geometry of motion, this transformation may be regarded as a **rotation** of points of the plane. Such a motion converts the region (3), for example, into the region (4), as indicated in Fig. 67. The concentric circles about the origin then become the lines of motion. Each ray is converted into another ray at an angular distance θ from it.

Let us now consider the general case where α is any complex number. Both ρ and θ may have any constant values. We have then a combination of the two special cases already considered; that is, the point is rotated about the origin through the angle θ while it is at the same time moved along the ray on which it is rotated; for, if we have

$$\alpha = \rho(\cos\theta + i\sin\theta),$$
$$z = r(\cos\phi + i\sin\phi),$$

then by multiplication we get

$$w = r\rho(\cos\overline{\theta + \phi} + i\sin\overline{\theta + \phi});$$

that is

$$mod\, w = r \cdot \rho, \qquad amp\, w = \theta + \phi. \tag{1}$$

The result of the transformation $w = \alpha z$ may be obtained, as we shall now show, by regarding each point as moving along a logarithmic spiral whose asymptotic point is at the origin. For this reason it is convenient to describe the given transformation as a **logarithmic spiral motion** about the origin.

Let

$$w = r'(\cos\phi' + i\sin\phi')$$

be the point into which the point

$$z = r(\cos\phi + i\sin\phi)$$

is mapped by means of the given transformation $w = \alpha z$. From (1) we then have

$$r' = r \cdot \rho, \qquad \phi' = \phi + \theta.$$

A logarithmic spiral passing through z may be written in the form

$$r = ce^{k\phi}.$$

In order that this same curve shall also pass through w, it is sufficient that

$$\begin{aligned} r' &= ce^{k\phi'} \\ &= ce^{k(\phi+\theta)} \\ &= ce^{k\phi}e^{k\theta} \\ &= re^{k\theta}; \end{aligned}$$

that is, it is sufficient that

$$r' = r \cdot \rho = re^{k\theta},$$

or

$$k = \frac{\log \rho}{\theta}.$$

Hence, if the logarithmic spiral whose equation is

$$r = ce^{\frac{\log \rho}{\theta}\phi} \tag{2}$$

passes through the point z it also passes through the point w into which z is mapped by the given transformation. Consequently, the given z-point may be regarded as passing into the corresponding w-point by a motion along this spiral.

As ρ and θ are both determined by α, it follows that α determines the particular system of spirals given in (2) along which any z-point may move into the corresponding w-point. The arbitrary constant c is the parameter of the system, and to each point z there corresponds one and only one value of c and hence one and only one logarithmic spiral of the system. The entire plane is filled by this system of curves, coiled up within each other.

37. The transformation $w = \alpha z + \beta$. The significance of this transformation may be most readily seen by regarding it as a combination of the two preceding transformations. Let a logarithmic spiral motion take place about the origin, and then let the result be translated by adding the number β. Analytically, this result is equivalent to introducing the auxiliary variable z', defined by the equation

$$z' = \alpha z, \tag{1}$$

and following this transformation by that of

$$w = z' + \beta. \tag{2}$$

The given transformation is therefore equivalent to a logarithmic spiral motion, that is a rotation and a stretching, followed by a translation.

The question naturally arises as to whether we may not reverse these two operations; namely, whether we may not take first the translation and then the logarithmic spiral motion. Analytically, the relation $w = \alpha z + \beta$ may be obtained by first introducing the auxiliary variable

$$z'' = z + \frac{\beta}{\alpha}, \tag{3}$$

and then putting

$$w = \alpha z''. \tag{4}$$

The amount of the rotation and stretching, that is the extent of the logarithmic spiral motion, given by (4) and (1), is determined by the complex constant α. Since the value of α is the same in both equations, the motion is the same. The translations defined by (3) and (2) are, however, different. Hence, we see that a logarithmic spiral motion and a translation are processes that can not be interchanged. As we have seen, the processes of rotation and stretching are on the other hand interchangeable processes.

Whenever we apply the transformation $w = \alpha z + \beta$, there is one point of the plane that remains unchanged; that is, there is one invariant point. We can readily locate this point, since in this case the z-point is identical with the corresponding w-point. Let us therefore write

$$z = \alpha z + \beta,$$

and from this equation determine the value of z.

Then, for $\alpha \neq 1$, we have as the invariant point

$$z = \frac{\beta}{1 - \alpha}.$$

The given transformation represents a logarithmic spiral motion about this invariant point. For, referring the points of the plane to the invariant point as the origin, that is putting

$$w = w' + \frac{\beta}{1 - \alpha}, \qquad z = z' + \frac{\beta}{1 - \alpha},$$

we have

$$w' + \frac{\beta}{1 - \alpha} = \alpha\left(z' + \frac{\beta}{1 - \alpha}\right) + \beta,$$

or

$$w' = \alpha z'. \tag{5}$$

This equation represents a logarithmic spiral motion about the new origin, that is about the invariant point

$$z = \frac{\beta}{1 - \alpha}.$$

As we have already seen, a logarithmic spiral motion converts a given configuration into a similar configuration. The amount of rotation that takes place in the logarithmic spiral motion represented by the transformation $w = \alpha z + \beta$ is given by the amplitude of α. The magnification that takes place in the elements of the configuration by means of this transformation is determined by $|\alpha| = \rho$; for, we have

$$|D_z w| = |D_z(\alpha z + \beta)| = |\alpha| = \rho.$$

Since all elements are magnified by the same amount and otherwise the configuration remains unchanged except in position, we may conclude that the general linear transformation

$$w = \alpha z + \beta$$

transforms the complex plane into itself in such a manner that any given configuration is converted into a similar configuration whose position is determined by α and the constant $\dfrac{\beta}{1 - \alpha}$, and whose relative size is determined by $|\alpha|$ alone.

Conversely, we may show that any transformation of the plane into itself which preserves the similarity of the figure is a linear transformation of the form $w = \alpha z + \beta$. Suppose that we have given two similar plane figures. Since a linear function of the form under discussion has two arbitrary constants, a function of this kind can always be found that will transform any two distinct points z_1, z_2 of the one configuration into any two given distinct points, say the points w_1, w_2, of the second configuration homologous respectively to z_1, z_2. For the determination of these two constants we have the two equations

$$w_1 = \alpha z_1 + \beta,$$
$$w_2 = \alpha z_2 + \beta,$$

whence, we obtain

$$\alpha = \frac{\begin{vmatrix} w_1 & 1 \\ w_2 & 1 \end{vmatrix}}{\begin{vmatrix} z_1 & 1 \\ z_2 & 1 \end{vmatrix}}, \quad \beta = \frac{\begin{vmatrix} z_1 & w_1 \\ z_2 & w_2 \end{vmatrix}}{\begin{vmatrix} z_1 & 1 \\ z_2 & 1 \end{vmatrix}}.$$

The functional relation that transforms the two points z_1, z_2 into the two points w_1, w_2 is therefore

$$w = \frac{\begin{vmatrix} w_1 & 1 \\ w_2 & 1 \end{vmatrix}}{\begin{vmatrix} z_1 & 1 \\ z_2 & 1 \end{vmatrix}} z + \frac{\begin{vmatrix} z_1 & w_1 \\ z_2 & w_2 \end{vmatrix}}{\begin{vmatrix} z_1 & 1 \\ z_2 & 1 \end{vmatrix}}. \tag{6}$$

The amount of rotation and stretching that takes place in this transformation is determined by

$$\frac{\begin{vmatrix} w_1 & 1 \\ w_2 & 1 \end{vmatrix}}{\begin{vmatrix} z_1 & 1 \\ z_2 & 1 \end{vmatrix}} = \frac{w_1 - w_2}{z_1 - z_2};$$

the rotation is given by the amplitude of this ratio, that is by *amp* $(w_1 - w_2)$ − *amp* $(z_1 - z_2)$, while the modulus of this quotient gives the ratio of magnification of the element $z_1 - z_2$.

In addition to this rotation and magnification, the transformation given by (6) involves a translation of the points of the complex plane. The amount and direction of this translation is determined by the quotient

$$\frac{\begin{vmatrix} z_1 & w_1 \\ z_2 & w_2 \end{vmatrix}}{\begin{vmatrix} z_1 & 1 \\ z_2 & 1 \end{vmatrix}}.$$

Since the two configurations are similar, the amount of rotation, stretching, and translation necessary to transform any element $z_1 - z_2$ into its corresponding element $w_1 - w_2$ will also transform any other element into its corresponding element. The required transformation is fully determined when the values of α and β are expressed in terms of known values, and consequently equation (6) gives the transformation sought.

If it is known that there exists a relation between w and z which is holomorphic in a certain portion of the complex plane and if by means of this functional relation a given configuration is transformed into one similar to it, then it is possible to show by a consideration of the derivative that this relation is linear. As already pointed out, the ratio of magnification that takes place in passing from the Z-plane to the W-plane by means of a transformation $w = f(z)$ is given by the modulus of $D_z w$, while the amplitude of this derivative gives the

rotation that takes place. In the case under consideration both the ratio of magnification and the rotation are constants for the various values of z and hence the derivative itself is constant, say equal to α. Writing

$$D_z w = \alpha,$$

we have upon integrating,

$$w = \alpha z + \beta,$$

where β is an arbitrary constant of integration. This constant of integration represents a translation in the plane, and by its proper selection the points of the one configuration finally go over into the corresponding points of the similar configuration, with which the desired conclusion is established.

38. The transformation $w = \frac{1}{z}\cdot$ If in the general linear fractional transformation, we put $\alpha = \delta = 0$, $\beta = \gamma = 1$, we have a very important special case, namely,

$$w = \frac{1}{z}\cdot$$

We shall now consider some of the properties of this transformation. If we write z in terms of polar coördinates, we have

$$z = \rho(\cos\theta + i\sin\theta).$$

Hence, we may write

$$w = \frac{1}{z} = \frac{1}{\rho(\cos\theta + i\sin\theta)}$$
$$= \frac{1}{\rho}\{\cos(-\theta) + i\sin(-\theta)\}.$$

Putting
$$w = \rho'(\cos\theta' + i\sin\theta'),$$
we have
$$\rho' = \frac{1}{\rho}, \quad \theta' = -\theta.$$

Geometrically, we may consider this transformation as made up of two parts. Let the point P (Fig. 68) represent any complex number z. Draw through P the line OP passing also through the origin. Upon OP find a point P' so that $OP' = \rho' = \frac{1}{\rho}\cdot$ The location of this point may be then considered as the first step in the geometrical interpretation of the given transformation. This operation is called **geometric inversion.** The second step consists in

rotating the point P' about the axis of reals until it again falls into the plane of the paper, that is through an angle of 180°. We shall call this process a **reflection** upon the axis of reals. The given transformation may be called a **reciprocation** and consists of a geometrical inversion followed by a reflection upon the axis of reals.

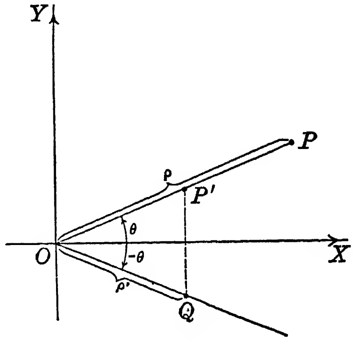

FIG. 68.

We shall first consider the properties of geometric inversion. This process is one that belongs to ordinary metrical geometry. If we draw about the origin a circle of unit radius, any point upon this circle will invert into itself, that is it remains invariant by the process of inversion. Every point within this unit circle is converted by this process into a point lying without it and *vice versa*. Every line drawn through the origin is converted into itself, except that the points are rearranged upon the line. The points very near to the origin are converted into points lying at a great distance and conversely. As we have already seen, it is convenient to regard the complex plane as closed at infinity, that is, as having a single point at infinity. This point at infinity inverts into the origin and *vice versa*.

To determine the character of the configurations into which configurations other than straight lines through the origin are inverted, we shall now turn to the analytic side of inversion. Suppose that by inversion z is changed into z'. (Since this process differs from the transformation $w = \dfrac{1}{z}$ in that the reflection upon the axis of reals is omitted,) we then have

$$
\begin{aligned}
z' &= \frac{1}{\rho}(\cos\theta + i\sin\theta) \\
&= \frac{\rho(\cos\theta + i\sin\theta)}{\rho^2} \\
&= \frac{x + iy}{x^2 + y^2}.
\end{aligned}
$$

Putting $z' = x' + iy'$, we have

$$x' + iy' = \frac{x + iy}{x^2 + y^2} = \frac{x}{x^2 + y^2} + i\frac{y}{x^2 + y^2},$$

or equating the real parts and the imaginary parts, we get

$$x' = \frac{x}{x^2 + y^2}, \qquad y' = \frac{y}{x^2 + y^2}. \tag{1}$$

Solving these equations for x and y, we have

$$x = \frac{x'}{x'^2 + y'^2}, \qquad y = \frac{y'}{x'^2 + y'^2}. \tag{2}$$

These are the values of x and y which, if substituted in the equation of a given curve, give the equation of the inverse curve.

EX. 1. Find the curve into which a straight line not passing through the origin is mapped by geometric inversion.

The equation of the given line is

$$Ax + By + C = 0, \qquad C \neq 0.$$

Substituting for x, y their values from (2) we obtain

$$\frac{Ax'}{x'^2 + y'^2} + \frac{By'}{x'^2 + y'^2} + C = 0,$$

$$C(x'^2 + y'^2) + Ax' + By' = 0.$$

This is the equation of a circle passing through the origin. The equation of the tangent to this circle at the origin is

$$Ax + By = 0,$$

which is a line parallel to the given line. Therefore, a system of parallel lines inverts into a system of circles having a common tangent at the origin. For $C = 0$, we have the special case of a line through the origin already discussed.

EX. 2. Find the curve into which a circle not passing through the origin is changed by inversion.

The equation of the given circle is of the form

$$x^2 + y^2 + 2gx + 2fy + C = 0.$$

Substituting the values of x, y from (2), we have

$$\frac{x'^2}{(x'^2 + y'^2)^2} + \frac{y'^2}{(x'^2 + y'^2)^2} + \frac{2gx'}{x'^2 + y'^2} + \frac{2fy'}{x'^2 + y'^2} + C = 0,$$

or
$$C(x'^2 + y'^2) + 2gx' + 2fy' + 1 = 0,$$

which is the equation of a circle not passing through the origin. In the special case where $C = 0$, we have a straight line not passing through the origin. If we now think of a straight line as a circle of infinite radius, we may then make the general statement that by geometric inversion every circle is converted into a circle.

The angle at which two curves cut each other is preserved by geometric inversion, but the direction of the angle is reversed; that is, the angle is measured in the opposite direction after inversion. We shall first show that this statement holds when one of the given curves is a straight line passing through the center of inversion. Let A, A', B, B' (Fig. 69) be two sets of points which are inverse with respect to O. Lines through A, A' and B, B' pass through O. Moreover, we have

$$\overline{OA} \cdot \overline{OA'} = \overline{OB} \cdot \overline{OB'} = 1.$$

The angle at O is common to the two triangles OAB and $OA'B'$, and as the sides of the common angle are proportional, the two triangles are similar. Consequently,

$$\angle OAB = \angle OB'A'. \qquad (3)$$

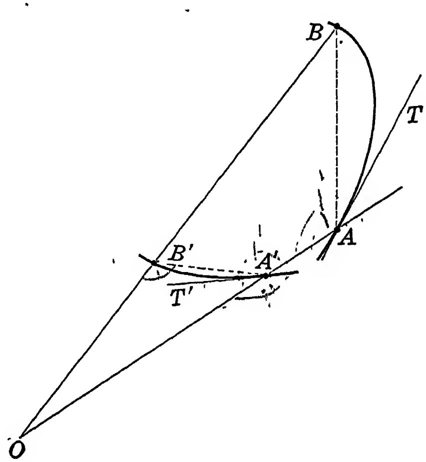

FIG. 69.

If now we think of the points A and B as situated on some curve, the points A' and B' will lie upon the inverse curve. Let the point B approach A as a limit. Then the point B' approaches the point A' along the corresponding curve. The lines AB and $A'B'$ become the tangents AT and $A'T'$ to the two curves at the corresponding points A and A', respectively. In the limit, therefore, the angle $OB'A'$ becomes the angle vertical to $AA'B'$ and hence equal to it. From (3) we then have in the limit

$$\angle OAT = \angle AA'T'.$$

The line OA is its own inverse since it passes through the center of inversion, and by hypothesis the curve $A'B'$ is the inverse of the curve AB. By inversion the angle that the tangent to the curve AB makes with OA, measured in a clockwise direction, namely $\angle A'AT$, is changed into the equal angle $\angle AA'T'$ made by the tangent to the inverse curve $A'B'$ with OA, measured in a counter-clockwise direction.

Suppose we now consider the case of any two curves intersecting in a point A. The inverse curves will intersect in a point A' which is the inverse of A. To extend the argument to this case, draw a straight line through A, A'. It will pass through the center of in-

version. Consider the angle made by the tangent to each curve with this line at the point of intersection. From the foregoing discussion this angle is preserved in magnitude but reversed in direction by inversion. By combination of these angles we have the desired result; that is, by geometric inversion angles are preserved in magnitude but reversed in direction.

Inversion is therefore a conformal transformation but with a reversion of any given angle. Reflection upon a straight line is likewise a process that involves a reversion of angles. As we have seen, the transformation $w = \frac{1}{z}$ is made up of a geometric inversion and a reflection upon the axis of reals. When we combine these two processes we have a process in which these two reversions annul each other. Hence, we can say that the transformation $w = \frac{1}{z}$ is conformal without reversion of angles.

We have thus far confined our discussion to geometric inversion with respect to the unit circle, because of the fact that inversion with respect to this circle is involved in the transformation $w = \frac{1}{z}$. This restriction, however, is not essential to the geometry of inversion. We may define inversion with respect to a circle of radius k by merely replacing the above condition $\rho\rho' = 1$ by the more general one $\rho\rho' = k^2$. To show that the same geometric properties hold for the general case suppose we think of the whole plane as being so expanded or contracted about the origin that the unit circle changes into the required circle of radius k. Any two corresponding points P, P' with respect to the unit circle become two corresponding points Q, Q' with respect to the required circle. If ρ_1, ρ_1' are the radii vectores of the points Q, Q', then we have

$$\begin{aligned} \rho_1 \rho_1' &= k\rho \cdot k\rho' \\ &= k^2\rho\rho' \\ &= k^2, \end{aligned}$$

as was required.

Two corresponding points with respect to a circle of inversion are called **conjugate points.** The conjugate of any particular point with respect to a given circle may be found geometrically as follows. Draw two tangents from the given point A to the given circle (Fig. 70). Join the given point with the center O of the circle. Connect the points of tangency B and C. The intersection of the chord BC

and the line OA gives the required point A'. For, from the figure, the triangle AOB is a right triangle, having a right angle at B, and hence we have

$$\overline{OB}^2 = \overline{A'O} \cdot \overline{AO},$$

showing the two points A and A' to be conjugate points.

The foregoing construction holds when the given point A lies without the circle of inversion. If the given point lies within the circle of inversion the conjugate point may be found as follows. Connect the given point A with the center O of the circle; through A draw a line perpendicular to AO, and at the points where this

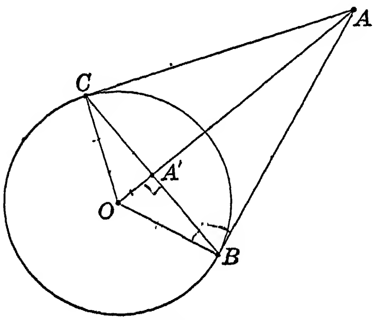

FIG. 70.

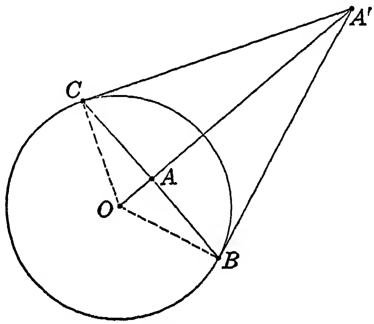

FIG. 71.

perpendicular intersects the circle draw tangents to the circle. The point of intersection of these tangents gives the required point A'. The proof is similar to that in the previous case.

The following theorems give additional important properties of inversion.

THEOREM I. *If a given circle cuts the circle of inversion in two points A and B, then its inverse cuts the circle of inversion in the same two points.*

The truth of this theorem is seen at once from the fact that the points of intersection of the given circle with the circle of inversion are points on the circle of inversion and therefore necessarily invert into themselves. It does not follow, of course, that the given circle as a whole inverts into itself.

THEOREM II. *If a given circle cuts the circle of inversion at a given angle, then its inverse cuts the circle of inversion at the same angle.*

The theorem follows from the fact that the magnitude of an angle is preserved by the process of inversion, and hence the angle at which the given circle cuts the circle of inversion remains unchanged in magnitude. It is, however, reversed in direction.

COROLLARY. *If a given circle cuts the circle of inversion at right angles, then the given circle is identical with its inverse and any straight line through the center of inversion cuts the given circle in two conjugate points.*

If the given circle cuts the circle of inversion at right angles, then by Theorem II its inverse also cuts the circle of inversion at the same angle, and since through two points on a circle but one orthogonal circle can be drawn, the given circle must be identical with its inverse, as the theorem requires. The only change that takes place in the given circle is that the portion of the circle without the circle of inversion becomes after inversion the portion within the circle of inversion. Since the given circle inverts into itself, it follows that any straight line passing through the origin cuts the given circle in conjugate points.

THEOREM III. *Given a pair of conjugate points with respect to a fixed circle. Any circle through these points inverts into itself with respect to the fixed circle and cuts that circle at right angles.*

One of the two conjugate points must lie within and the other without the circle of inversion. Consequently, the given circle cuts the fixed circle and the two points of intersection invert into themselves. These points of intersection and the two given conjugate points make together four points that the given circle and the inverted circle have in common. Hence, the two circles must coincide. By Theorem II the given circle and the inverted circle cut the circle of inversion at the same angle but reversed in direction. But as the inverted circle is identical with the given circle each must then cut the circle of inversion at right angles.

THEOREM IV. *Given a system of circles such that each circle passes through two given points and intersects a fixed circle at right angles. The two given points of intersection of the system of circles are then conjugate points with respect to the fixed circle.*

Let the circles of the system be inverted with respect to the given fixed circle M. By the corollary to Theorem II, each circle of the system inverts into itself. It is sufficient for our purpose to con-

sider two circles C_1, C_2 of the system. The point P of intersection lies on both C_1, and C_2. After inversion with respect to M, the point P must go into a point within M which likewise lies upon both C_1 and C_2. It must, therefore, invert into the second point of intersection of these two circles, namely P'. Hence the theorem.

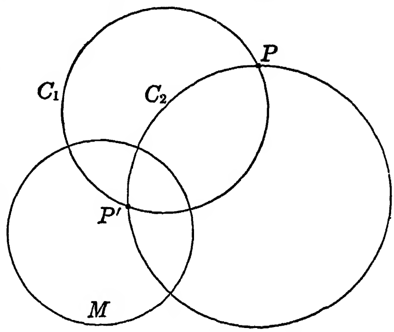

Fig. 72.

Theorem V. *Given two conjugate points with respect to a given circle. If the circle is inverted with respect to a fixed circle, the given conjugate points invert into conjugate points with respect to the inverted circle.*

Let M be the given circle and P and P' two conjugate points with respect to it. Suppose the circle M inverts into the circle M' with respect to the fixed circle C, and the points P, P' invert into Q, Q', respectively. It is required to show that Q, Q' are conjugate points with respect to M'. Draw any two circles through the conjugate points P, P'; these circles cut the given circle M at right angles. These angles are preserved by inversion. Hence, the circles through the given conjugate points and cutting M at right angles invert into circles cutting M' at right angles. Since the inverse points of P, P', namely the points Q, Q', must lie at the respective intersections of these inverted circles, it follows from Theorem IV that the points Q, Q' are conjugate points with respect to the circle M'. With this our theorem is demonstrated.

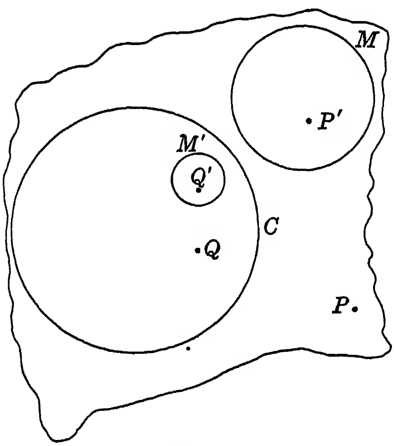

Fig. 73.

39. General properties of the transformation $w = \dfrac{\alpha z + \beta}{\gamma z + \delta}$. We shall now consider the general case of a linear fractional transfor-

mation. We impose the condition upon the four constants α, β, γ, δ, that

$$\begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} = \alpha\delta - \beta\gamma \neq 0. \tag{1}$$

If this determinant were equal to zero, we should have $\frac{\alpha}{\gamma} = \frac{\beta}{\delta}$ and the given relation between w and z would then reduce to

$$w = \frac{\alpha}{\gamma},$$

and all points in the Z-plane would correspond to the same point $\frac{\alpha}{\gamma}$ in the W-plane. By imposing the condition (1), we are able to set aside this trivial case.

The general linear fractional relation may be decomposed into the three following special cases, namely:

(1) $z' = z + \dfrac{\delta}{\gamma}$,

(2) $z'' = \dfrac{1}{z'}$,

(3) $w = \dfrac{\beta\gamma - \alpha\delta}{\gamma^2} z'' + \dfrac{\alpha}{\gamma}$.

This statement can be easily verified by making the substitutions indicated and thus obtaining the general linear fractional relation between w and z. Geometrically, we may then consider the general linear transformation as made up of the following:

(1) *a translation,*

(2) *a geometric inversion followed by a reflection on the axis of reals,*

(3) *a rotation and a stretching followed by a translation; or what is the same thing, a logarithmic spiral motion about the point left invariant by the third of the foregoing transformations.*

As we have already considered each of these operations, we can now formulate some of the general properties of a linear fractional transformation. Among these properties are:

THEOREM I. *Conjugate points with respect to a given circle are transformed by the general linear fractional transformation into conjugate points with respect to the transformed circle.*

We have seen (Theorem V, Art. 38) that conjugate points with respect to a given circle remain conjugate points by inversion. Since reflection upon the axis of reals does not disturb the relative position of points of a given configuration except to reverse the direction of the angles, we may conclude that the theorem holds for the special transformation $w = \frac{1}{z}$; that is, it holds for the transformation (2) given above. It also holds for the transformations (1) and (3) since by both these transformations the similarity of the configuration is preserved. As the general linear transformation is decomposable into these three special transformations, for each of which conjugate points remain conjugate points, the theorem follows as stated.

THEOREM II. *Any given configuration is mapped conformally, without reversion of angles, by means of a linear fractional transformation.*

This theorem follows from the fact that each of the three simple transformations into which the general linear fractional transformation may be decomposed is such that the conclusions stated in the theorem hold.

Since w is holomorphic for all values of z in the finite region except for $z = -\frac{\delta}{\gamma}$, this same result may be obtained independently by the consideration of $D_z w$. We have

$$D_z w = \frac{\alpha\delta - \beta\gamma}{(\gamma z + \delta)^2}.$$

But by hypothesis

$$\alpha\delta - \beta\gamma \neq 0.$$

Hence, by the theorem of Art. 27, the desired result follows.

THEOREM III. *By the general linear fractional transformation, circles are converted into circles.*

It is here understood that a straight line is to be considered as a circle of infinite radius. The truth of the theorem follows from the fact that it holds for each of the three special transformations into which the general linear transformation may be decomposed.

THEOREM IV. *The general linear fractional transformation leaves two points in the complex plane invariant.*

To establish this theorem, we proceed as follows. If any point

z of the complex plane is transformed into itself by means of a linear fractional transformation, then we must have

$$z = \frac{\alpha z + \beta}{\gamma z + \delta},$$

that is

$$\gamma z^2 + (\delta - \alpha) z - \beta = 0. \tag{2}$$

This equation is a quadratic and has therefore two roots, namely:

$$z_1 = \frac{(\alpha - \delta) + \sqrt{(\alpha - \delta)^2 + 4\beta\gamma}}{2\gamma}, \quad z_2 = \frac{(\alpha - \delta) - \sqrt{(\alpha - \delta)^2 + 4\beta\gamma}}{2\gamma}. \tag{3}$$

The two points z_1, z_2 remain unchanged by the general linear fractional transformation, since each is transformed into itself.

These invariant points may be finite and distinct, finite and coincident, one finite and the other infinite, or finally, both may be infinite. The analytic conditions for these various cases may be expressed in terms of the coefficients of (2). If the discriminant vanishes, that is if

$$(\alpha - \delta)^2 + 4\beta\gamma = 0,$$

the two roots of (1), that is the two invariant points, are coincident. If in addition we have $\gamma = 0$, it will be seen from (1) that both roots of (1) become infinite; that is, both invariant points coincide at the point infinity. If $\gamma \neq 0$ the two points z_1, z_2 lie in the finite region of the plane.

If we have

$$(\alpha - \delta)^2 + 4\beta\gamma \neq 0,$$

the roots of (1), that is the invariant points, are distinct. If in addition $\gamma = 0$, one of the roots of (2) becomes infinite and hence one of the invariant points is at infinity. It will be observed that when $\gamma = 0$ the linear fractional transformation reduces to the general linear transformation.

The general linear fractional transformation contains four constants; but as we may divide both numerator and denominator by one of these without affecting the transformation, we have only three independent constants. We may state the following theorem.

THEOREM V. *There is always one and only one linear fractional transformation that transforms any three distinct points into three given distinct points.*

Let z_1, z_2, z_3 be the three distinct points that are to be transformed into the given distinct points w_1, w_2, w_3. We must then have the three relations

$$w_k = \frac{\alpha z_k + \beta}{\gamma z_k + \delta}, \qquad k = 1, 2, 3;$$

that is,

$$\gamma w_k z_k + \delta w_k - \alpha z_k - \beta = 0. \tag{4}$$

We have given three linear homogeneous equations in the four unknowns α, β, γ, δ. The condition that these equations have one and only one solution other than $\alpha = \beta = \gamma = \delta = 0$ is that the matrix of the coefficients, or its equivalent matrix,

$$\left\|\begin{matrix} w_1z_1 & w_1 & z_1 & 1 \\ w_2z_2 & w_2 & z_2 & 1 \\ w_3z_3 & w_3 & z_3 & 1 \end{matrix}\right\| \tag{5}$$

shall be of rank three;* that is, that not all of the determinants formed from this matrix by dropping one column shall vanish.

We shall show that this condition is satisfied by showing that the two determinants

$$\Delta_1 = \begin{vmatrix} w_1 & z_1 & 1 \\ w_2 & z_2 & 1 \\ w_3 & z_3 & 1 \end{vmatrix}, \qquad \Delta_2 = \begin{vmatrix} w_1z_1 & z_1 & 1 \\ w_2z_2 & z_2 & 1 \\ w_3z_3 & z_3 & 1 \end{vmatrix}$$

can not vanish simultaneously. Expanding each determinant in terms of the elements of the first column, we have

$$\begin{aligned} \Delta_1 &= w_1(z_2 - z_3) - w_2(z_1 - z_3) + w_3(z_1 - z_2), \\ \Delta_2 &= w_1z_1(z_2 - z_3) - w_2z_2(z_1 - z_3) + w_3z_3(z_1 - z_2). \end{aligned}$$

Multiplying the first of these identities by z_1 and subtracting the second from that result we get

$$\begin{aligned} z_1\Delta_1 - \Delta_2 &= -w_2(z_1 - z_2)(z_1 - z_3) + w_3(z_1 - z_2)(z_1 - z_3) \\ &= (w_3 - w_2)(z_1 - z_2)(z_1 - z_3). \end{aligned} \tag{6}$$

Since the points z_1, z_2, z_3 and w_1, w_2, w_3 are distinct, it follows that (6) can not vanish. Hence, we have

$$z_1\Delta_1 - \Delta_2 \neq 0,$$

and consequently the two determinants Δ_1, Δ_2 can not vanish simultaneously.

* See Bôcher, *Introduction to Higher Algebra*, Art. 17.

Since the equations (1) have one and only one solution other than $\alpha = \beta = \gamma = \delta = 0$, it follows that any three ratios of these unknowns are uniquely determined. Consequently there is one and only one transformation of the required type which transforms the three distinct points z_1, z_2, z_3 into three distinct points w_1, w_2, w_3. Hence the theorem.

Remembering that three distinct points definitely determine a circle we may now say that any circle can be transformed into any other circle, or into itself, by means of a linear fractional transformation. Since the three points upon the given circle can be selected in an infinite number of ways, it follows that the required transformation can be made in an infinite number of ways.

If it is desired to transform four points into four points, we must have an additional condition satisfied. If we have given any four points z_1, z_2, z_3, z_4 the ratio

$$\frac{z_1 - z_2}{z_3 - z_2} : \frac{z_1 - z_4}{z_3 - z_4}$$

is called the **anharmonic ratio or cross-ratio** of these four points.

The following theorem gives the condition which must be satisfied in order that any four distinct points z_1, z_2, z_3, z_4 may be transformed into four distinct points w_1, w_2, w_3, w_4 by a linear fractional transformation.

THEOREM VI. *The necessary and sufficient condition that any four distinct points of the complex plane may be transformed by a linear fractional transformation into any other four distinct points of the plane is that the anharmonic ratio of the two sets of points is the same.*

Let the four given points be z_1, z_2, z_3, z_4 and let it be required to transform these points into the four distinct points w_1, w_2, w_3, w_4. If the four given z-points are transformed by a linear fractional transformation into the four given w-points, then we must have the four relations

$$w_k = \frac{\alpha z_k + \beta}{\gamma z_k + \delta}, \qquad k = 1, 2, 3, 4, \tag{7}$$

or

$$\gamma w_k z_k + \delta w_k - \alpha z_k - \beta = 0.$$

The necessary and sufficient condition that these four equations have

a solution other than $\alpha = \beta = \gamma = \delta = 0$ is that the determinant of the coefficients shall vanish; that is, that we have *

$$\begin{vmatrix} w_1z_1 & w_1 & -z_1 & -1 \\ w_2z_2 & w_2 & -z_2 & -1 \\ w_3z_3 & w_3 & -z_3 & -1 \\ w_4z_4 & w_4 & -z_4 & -1 \end{vmatrix} = \begin{vmatrix} w_1z_1 & w_1 & z_1 & 1 \\ w_2z_2 & w_2 & z_2 & 1 \\ w_3z_3 & w_3 & z_3 & 1 \\ w_4z_4 & w_4 & z_4 & 1 \end{vmatrix} = 0.$$

Expanding this determinant in terms of the last two columns by Laplace's development, we have

$$\begin{vmatrix} z_1 & 1 \\ z_2 & 1 \end{vmatrix} \cdot \begin{vmatrix} z_3 & 1 \\ z_4 & 1 \end{vmatrix} w_3w_4 - \begin{vmatrix} z_1 & 1 \\ z_3 & 1 \end{vmatrix} \cdot \begin{vmatrix} z_2 & 1 \\ z_4 & 1 \end{vmatrix} w_2w_4 + \begin{vmatrix} z_1 & 1 \\ z_4 & 1 \end{vmatrix} \cdot \begin{vmatrix} z_2 & 1 \\ z_3 & 1 \end{vmatrix} w_2w_3$$
$$+ \begin{vmatrix} z_3 & 1 \\ z_4 & 1 \end{vmatrix} \cdot \begin{vmatrix} z_1 & 1 \\ z_2 & 1 \end{vmatrix} w_1w_2 - \begin{vmatrix} z_2 & 1 \\ z_4 & 1 \end{vmatrix} \cdot \begin{vmatrix} z_1 & 1 \\ z_3 & 1 \end{vmatrix} w_1w_3 + \begin{vmatrix} z_2 & 1 \\ z_3 & 1 \end{vmatrix} \cdot \begin{vmatrix} z_1 & 1 \\ z_4 & 1 \end{vmatrix} w_1w_4.$$

Making use of the identity

$$\begin{vmatrix} z_1 & 1 \\ z_2 & 1 \end{vmatrix} \cdot \begin{vmatrix} z_3 & 1 \\ z_4 & 1 \end{vmatrix} - \begin{vmatrix} z_1 & 1 \\ z_3 & 1 \end{vmatrix} \cdot \begin{vmatrix} z_2 & 1 \\ z_4 & 1 \end{vmatrix} + \begin{vmatrix} z_1 & 1 \\ z_4 & 1 \end{vmatrix} \cdot \begin{vmatrix} z_2 & 1 \\ z_3 & 1 \end{vmatrix} = 0,$$

we may write the foregoing relation in the form

$$\begin{aligned} &(z_1 - z_2)(z_3 - z_4)\,\{(w_3w_4 + w_1w_2) - (w_2w_4 + w_1w_3)\} \\ &\quad + (z_1 - z_4)(z_2 - z_3)\,\{(w_2w_3 + w_1w_4) - (w_2w_4 + w_1w_3)\} \\ &= -(z_1 - z_2)(z_3 - z_4)(w_1 - w_4)(w_3 - w_2) \\ &\quad + (z_1 - z_4)(z_3 - z_2)(w_1 - w_2)(w_3 - w_4) = 0, \end{aligned}$$

whence we get

$$\frac{z_1 - z_2}{z_3 - z_2} : \frac{z_1 - z_4}{z_3 - z_4} = \frac{w_1 - w_2}{w_3 - w_2} : \frac{w_1 - w_4}{w_3 - w_4};$$

that is, the anharmonic ratio of the four points z_1, z_2, z_3, z_4 is the same as the anharmonic ratio of the four points w_1, w_2, w_3, w_4. As this result presents the necessary and sufficient condition that the equations (4) have a solution other than $\alpha = \beta = \gamma = \delta = 0$, it follows that this result also gives the necessary and sufficient condition that the one set of four points may be mapped by a linear fractional transformation into the other set. Consequently, the theorem follows as stated.

It may be remarked that as a consequence of the foregoing theorem a linear fractional transformation has the property that it leaves the anharmonic ratio of any four points invariant.

If the order in which the four given points are taken is changed

* See Bôcher, *Introduction to Higher Algebra*, Art. 17, Theorem 3, Cor. 2.

then the anharmonic ratio may be changed. Of the twenty-four ways in which four points may be selected, only six give distinct anharmonic ratios. If we denote any one of these ratios by λ, then the six are given by *

$$\lambda, \quad \frac{1}{\lambda}, \quad 1-\lambda, \quad \frac{1}{1-\lambda}, \quad \frac{\lambda}{\lambda-1}, \quad \frac{\lambda-1}{\lambda}.$$

That λ is in general a complex number follows from the fact that it is defined as the ratio of such numbers. It may therefore be represented as a point in the complex plane. Since any two of these six anharmonic ratios are linearly related, the geometric interpretation of these relations furnishes an interesting exercise in the application of the principles developed in this chapter.

If λ describes a circle in the complex plane, then the points representing respectively the various ratios likewise describe circles. Moreover, if λ is represented by points within a given region bounded by a circle, it follows that the other ratios are represented by points within regions bounded by circles. It is possible to so choose the region for λ that the entire complex plane shall be filled by the regions of the six ratios without overlapping. The relative position of these regions may be found as follows. With $x = 0$ and $x = 1$ as centers describe two unit circles (Fig. 74). These circles intersect at the points whose coördinates x, y satisfy the two equations

$$x^2 + y^2 = 1,$$
$$(x-1)^2 + y^2 = 1.$$

By solving these equations, we have

$$x = \tfrac{1}{2}, \qquad y = \pm \tfrac{1}{2}\sqrt{3}.$$

The points of intersection are, therefore, $-\omega^2$ and $-\omega$, where

$$\omega = \frac{-1 + i\sqrt{3}}{2}$$

is one of the cube roots of unity. If λ takes the values in the unshaded region $(0, \tfrac{1}{2}, -\omega^2)$, Fig. 74, then $\dfrac{1}{\lambda}$ is confined to the region found by inverting this region with respect to the center 0 and reflecting the result upon the real axis. The numbers λ and $1-\lambda$ are symmetrical with respect to the point $\dfrac{1}{2}\cdot$ The region for $\dfrac{1}{1-\lambda}$ may be obtained

* See Scott, *Modern Analytical Geometry*, p. 37.

from that for $1 - \lambda$ by inverting this region with respect to the circle about the origin as a center and reflecting the result upon the axis of reals. From the region for $\frac{1}{\lambda}$ we may find the region for $\frac{\lambda - 1}{\lambda}$, be-

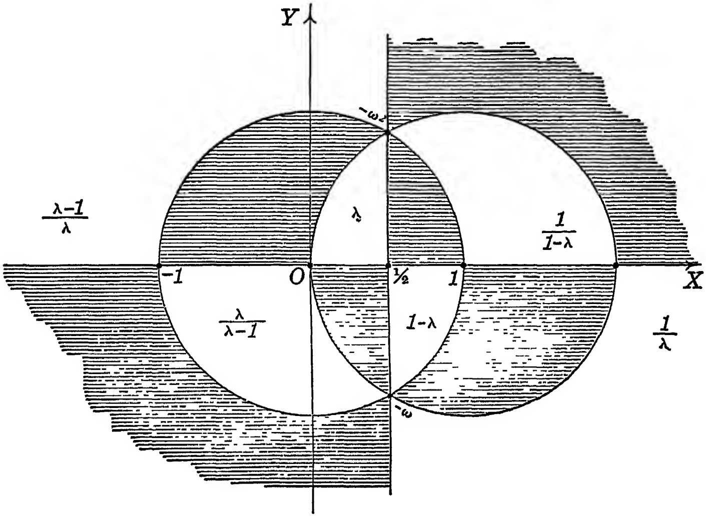

FIG. 74.

cause these two numbers are symmetric with respect to $\frac{1}{2}$. In this way we find the regions described by the various complex numbers as shown in the figure. If λ is represented by the points of a shaded region, then the points representing the other five anharmonic ratios are confined to the shaded regions.

There are two important special cases of anharmonic ratios. One of these cases is obtained if λ has such values that $\lambda = \frac{1}{\lambda}$ and hence $\lambda = \pm 1$. For $\lambda = -1$, the four points are said to be **harmonic.** The six ratios are then coincident in pairs.

When λ is a complex number, as in the present discussion, it is possible for three of the anharmonic ratios to be equal. For example, it will be seen from the figure that the three ratios λ, $\frac{1}{1 - \lambda}$, $\frac{\lambda - 1}{\lambda}$ may become equal at the common point,

$$-\omega^2 = \frac{1 + i\sqrt{3}}{2}.$$

The reciprocals of these values, that is $\frac{1}{\lambda}$, $1 - \lambda$, $\frac{\lambda}{\lambda - 1}$, then become equal at

$$-\omega = \frac{1 - i\sqrt{3}}{2}.$$

This equality leads to the second special case of anharmonic ratios; for, putting

$$\lambda = \frac{1}{1-\lambda} = \frac{\lambda - 1}{\lambda},$$

we have

$$\lambda^2 - \lambda + 1 = 0,$$

whence

$$\lambda = \frac{1 \pm i\sqrt{3}}{2},$$

which are the two imaginary cube roots of -1. When λ has either of these values the four points are said to be **equianharmonic.***

If the variables and constants involved in a linear fractional transformation are all real, the property that anharmonic ratios are preserved is commonly spoken of as a projective property; in fact this property may be made the basis of projective geometry. The relation between anharmonic ratios and linear fractional transformation, as established in Theorems V and VI, suggests the extension of projective geometry to the field of complex numbers. In the one case the single variable x takes the totality of real values and the ideal number ∞, represented by the points on a straight line including the point at infinity. As a result, we have the projective geometry of a straight line. In the other case the single variable z takes the totality of complex numbers and the ideal number ∞. Since but a single variable is involved this aggregate is sometimes spoken of in projective geometry as the complex line. This extension of projective geometry to the realm of complex numbers leads to the consideration of the theory of chains,† but, as no use will be made of this theory in the present volume, it will not be considered here.

In this connection it is also of interest to point out the general relation between the totality of linear fractional transformations and the theory of groups. We have for example the following theorem.

* For a more extended discussion of these cases see Harkness and Morley, *Treatise on the Theory of Functions*, p. 21, et seq.

† For a discussion of this subject, see J. W. Young, *Annals of Math.*, Vol. II, pp. 33–48.

THEOREM VII. *The system of linear fractional transformations possesses the group property.*

The statement contained in the theorem involves the condition that if a linear fractional expression in one variable is subjected to a linear fractional transformation, the resulting expression is a linear fractional expression. Given the relation

$$w = \frac{\alpha_1 z' + \beta_1}{\gamma_1 z' + \delta_1}, \qquad \begin{vmatrix} \alpha_1 & \beta_1 \\ \gamma_1 & \delta_1 \end{vmatrix} \neq 0.$$

Suppose z' is associated with z by the relation

$$z' = \frac{\alpha_2 z + \beta_2}{\gamma_2 z + \delta_2}, \qquad \begin{vmatrix} \alpha_2 & \beta_2 \\ \gamma_2 & \delta_2 \end{vmatrix} \neq 0.$$

The theorem requires that w be expressed as a linear fractional function of z, where the determinant of the coefficients is also different from zero. We have

$$\begin{aligned} w &= \frac{\alpha_1 \dfrac{\alpha_2 z + \beta_2}{\gamma_2 z + \delta_2} + \beta_1}{\gamma_1 \dfrac{\alpha_2 z + \beta_2}{\gamma_2 z + \delta_2} + \delta_1} \\ &= \frac{(\alpha_1\alpha_2 + \beta_1\gamma_2) z + (\alpha_1\beta_2 + \beta_1\delta_2)}{(\gamma_1\alpha_2 + \delta_1\gamma_2) z + (\gamma_1\beta_2 + \delta_1\delta_2)}, \end{aligned}$$

which is a linear fractional expression in z. The determinant of the coefficients is different from zero; for, we have

$$\begin{vmatrix} \alpha_1\alpha_2 + \beta_1\gamma_2 & \alpha_1\beta_2 + \beta_1\delta_2 \\ \gamma_1\alpha_2 + \delta_1\gamma_2 & \gamma_1\beta_2 + \delta_1\delta_2 \end{vmatrix} = \begin{vmatrix} \alpha_1 & \beta_1 \\ \gamma_1 & \delta_1 \end{vmatrix} \cdot \begin{vmatrix} \alpha_2 & \beta_2 \\ \gamma_2 & \delta_2 \end{vmatrix},$$

and each determinant in the second member of this equation is different from zero by hypothesis.

The system of linear fractional transformations possesses the other characteristic properties of a group,* and the relation to the theory of groups is at once established. If α, β, γ, δ are integers such that

$$\begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} = 1,$$

then the transformation

$$w = \frac{\alpha z + \beta}{\gamma z + \delta}$$

* See Bôcher, *Introduction to Higher Algebra*, p. 82.

defines the modular group.* Many of the properties of linear fractional transformations that have been discussed follow also as applications of group theory.†

40. Stereographic projection. Since complex numbers are of the form $z = x + iy$, where x and y may vary independently of each other, two degrees of freedom are necessary for the geometric element used to interpret them and the plane naturally suggests itself for that purpose. Thus far we have restricted ourselves to this mode of representation. There are other ways, however, of representing complex numbers and other surfaces than the plane have been made use of in this connection. It is frequently convenient to employ the sphere for this purpose. In order to do so, it must be possible to establish in some way a one-to-one correspondence between the points of a plane and those upon the sphere.

The desired result may be accomplished by assuming the complex plane as before and supposing that we have a sphere tangent to this plane at the origin. We shall refer to the point of tangency as the south pole of the sphere, while the opposite pole will be spoken of as the north pole. If we now take the north pole O' as the center of projection we can project in a definite manner every point of the plane upon the sphere. Thus in Fig. 75 the point P in the complex plane corresponds to the point P' of the sphere. In this way there corresponds to each point of the plane a definite point of the sphere, and conversely. This method of mapping the complex plane upon the sphere is called **stereographic projection.**

Since there is a one-to-one correspondence between the points of the complex plane and those of the sphere the values of z and of $w = f(z)$ may be uniquely represented upon the sphere, which we shall refer to as the complex sphere. For example, if z describes a continuous curve in a region of the complex plane in which $w = f(z)$ is holomorphic, then w likewise describes a continuous curve. The projection of these two curves upon the sphere gives the interpretation upon that surface of the relation between w and z.

The point at infinity in the complex plane projects into the north pole of the sphere. Hence, to examine the nature of a function for values of the variable in the neighborhood of the point at infinity, it may often be convenient to represent both w and z upon the complex sphere and inquire into the behavior of w as z takes values in the

* See Forsyth, *Theory of Functions*, 2d Ed., pp. 680, 681.

† See Kowalewski, *Komplexen Veranderlichen und ihre Funktionen*, pp. 30–59.

neighborhood of the north pole. The same result, of course, could be obtained analytically. Corresponding to the coördinates x, y of a point in the plane, we may determine the location of a point upon the sphere by means of two coördinates θ, ϕ, one measured along

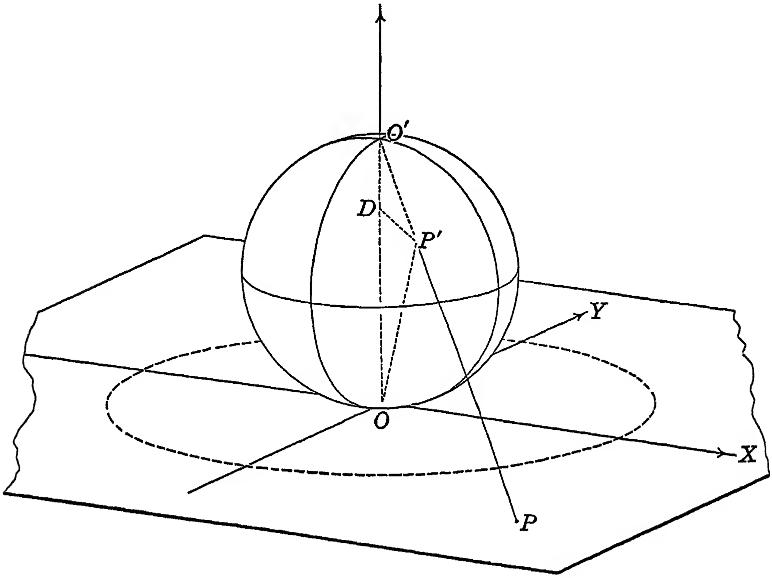

FIG. 75.

some standard meridian and the other along the equator. Such a system of coördinates is a familiar one in the location of a point upon the earth's surface by means of its longitude and latitude. Any given curve can be mapped from the plane upon the sphere by means of the analytic relation between x, y and the coördinates of the corresponding point on the sphere, and the transformed function thus obtained can be studied for values of θ, ϕ in the neighborhood of the north pole. A closed curve upon the sphere divides the surface of the sphere into two parts. This curve may be regarded as the boundary of either of these regions to suit our convenience.

It is desirable to examine somewhat more closely into the effect of stereographic projection upon the character of a configuration. First of all, suppose we have a pencil of rays passing through the origin and lying in the complex plane. Each of these rays projects into a great circle passing through O and O'. They become meridians upon the sphere and one such meridian passes through each point

upon the sphere. As a special case the axis of reals projects into a meridian of reals and the axis of imaginaries projects into a meridian of imaginaries cutting the meridian of reals at right angles.

If we have a system of concentric circles in the plane having the origin as center, they constitute the orthogonal system to the pencil of rays just mentioned. These circles go over into the orthogonal system of circles on the sphere, namely, the parallels of latitude. One of these circles projects into the equator of the sphere. This circle may be conveniently selected as the unit circle in the plane. All concentric circles lying within this unit circle will become parallels of latitude in the southern hemisphere while those lying outside of this unit circle pass over into parallels of latitude in the northern hemisphere.

In order to determine the character of a configuration on the sphere and its relation to the corresponding configuration in the plane, we shall now deduce the equations of transformation by means of which the cartesian space coördinates of any point upon the sphere can be expressed in terms of the cartesian coördinates of the corresponding point in the plane. Let ξ, η, ζ denote the coördinates of a point on the sphere. Let the ξ-axis and the η-axis coincide respectively with the axis of reals and the axis of imaginaries of the plane. Let the ζ-axis be perpendicular to the complex plane. Suppose the radius of the given sphere to be $\frac{1}{2}$. The equation of the sphere is

$$\xi^2 + \eta^2 + (\zeta - \tfrac{1}{2})^2 = \tfrac{1}{4}, \tag{1}$$

or

$$\xi^2 + \eta^2 + \zeta(\zeta - 1) = 0. \tag{2}$$

If we now denote by x, y the coördinates of any point P in the plane, the coördinates ξ, η, ζ of the projection P' upon the sphere of the point P are readily found in terms of x, y. From Fig. 75 we have

$$\overline{OP}^2 = x^2 + y^2, \tag{3}$$

$$\overline{O'P}^2 = \overline{O'O}^2 + \overline{OP}^2$$
$$= x^2 + y^2 + 1, \tag{4}$$

$$\overline{DP'}^2 = \xi^2 + \eta^2, \tag{5}$$

where DP' is drawn parallel to OP. The triangles OPO' and $DP'O'$ are similar, and consequently we have

$$\frac{\overline{OP}}{\overline{O'P}} = \frac{\overline{DP'}}{\overline{O'P'}}.$$

By use of (3), (4) and (5), we obtain

$$\frac{x^2 + y^2}{x^2 + y^2 + 1} = \frac{\xi^2 + \eta^2}{\overline{O'P'}^2}. \tag{6}$$

As $OP'O'$ is a right triangle, we have

$$\overline{O'P'}^2 = \overline{DO'} \cdot \overline{OO'} = \overline{DO'} \cdot 1.$$

Since $\overline{DO'} = 1 - \zeta$, we have

$$\overline{O'P'}^2 = 1 - \zeta. \tag{7}$$

From (6) we have then

$$\frac{x^2 + y^2}{x^2 + y^2 + 1} = \frac{\xi^2 + \eta^2}{1 - \zeta}. \tag{8}$$

We have also

$$\overline{DP'}^2 = \overline{DO'} \cdot \overline{OD},$$

from which we have

$$\overline{OD} = \frac{\overline{DP'}^2}{\overline{DO'}},$$

or

$$\zeta = \frac{\xi^2 + \eta^2}{1 - \zeta} = \frac{x^2 + y^2}{x^2 + y^2 + 1}. \tag{9}$$

Finally, we have

$$\frac{\xi}{\eta} = \frac{x}{y}, \tag{10}$$

or

$$\xi = \frac{x}{y}\eta. \tag{11}$$

By use of (9) and (11) the equation (2) of the given sphere may now be written

$$\left(1 + \frac{x^2}{y^2}\right)\eta^2 + \frac{x^2 + y^2}{x^2 + y^2 + 1} \cdot \frac{-1}{x^2 + y^2 + 1} = 0,$$

or

$$\eta^2 = \frac{x^2 + y^2}{(x^2 + y^2 + 1)^2} \cdot \frac{y^2}{x^2 + y^2} = \left\{\frac{y}{x^2 + y^2 + 1}\right\}^2,$$

whence

$$\eta = \frac{y}{x^2 + y^2 + 1}.$$

From (11) we get

$$\xi = \frac{x}{y} \cdot \frac{y}{x^2 + y^2 + 1} = \frac{x}{x^2 + y^2 + 1}.$$

Hence, the general relations between ξ, η, ζ and x, y are

$$\xi = \frac{x}{x^2 + y^2 + 1}, \qquad \eta = \frac{y}{x^2 + y^2 + 1}, \qquad \zeta = \frac{x^2 + y^2}{x^2 + y^2 + 1}. \tag{12}$$

From these equations we obtain

$$x = \frac{\xi}{1-\zeta}, \qquad y = \frac{\eta}{1-\zeta}, \qquad x^2 + y^2 = \frac{\zeta}{1-\zeta}. \tag{13}$$

Ex. 1. Find the stereographic projection of a straight line.

The equation of the given line is of the form

$$Ax + By + C = 0. \tag{14}$$

Substituting from (13) the values of x, y, we have

$$\frac{A\xi}{1-\zeta} + \frac{B\eta}{1-\zeta} + C = 0, \tag{15}$$

or

$$A\xi + B\eta + C(1-\zeta) = 0. \tag{16}$$

This equation is that of a plane passing through the north pole of the sphere. The curve of intersection of the plane and the sphere is therefore a circle passing

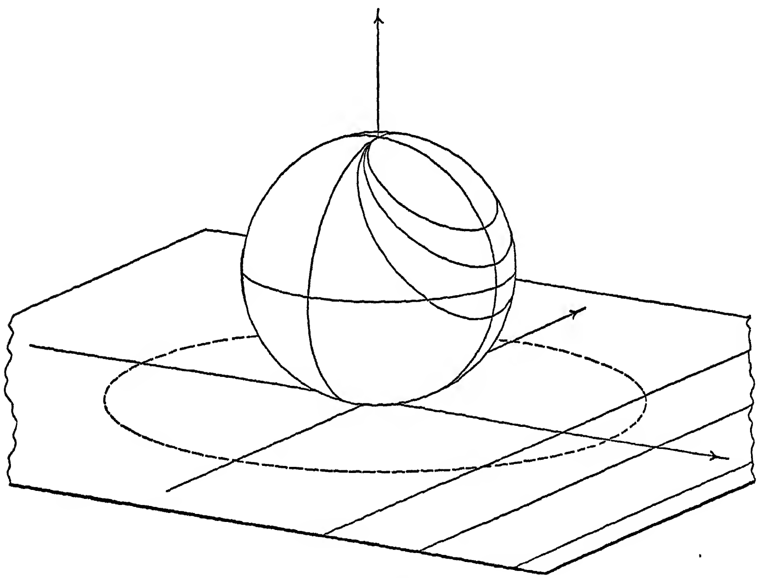

FIG. 76.

through the north pole. Hence every line in the plane projects into a circle upon the sphere passing through the north pole. Any line parallel, say, to the axis of imaginaries (Fig. 76) goes over into a circle through O' tangent to the great circle into which the axis of imaginaries projects, but lying wholly in one of the hemispheres into which that great circle divides the sphere. None of these lines, however, other than the one through the origin, projects into a great circle. A system of straight lines parallel to the axis of reals projects into a system of circles likewise passing through O' but perpendicular to the former system, and all of these circles on the sphere are tangent at O' to the great circle into which the axis of reals projects.

EX. 2. Discuss the stereographic projection of a circle in the complex plane whose equation is

$$x^2 + y^2 + 2gx + 2fy + c = 0. \tag{17}$$

Substituting from (13) the values of x, y, and $x^2 + y^2$ we have

$$\frac{\zeta}{1-\zeta} + 2g\frac{\xi}{1-\zeta} + 2f\frac{\eta}{1-\zeta} + c = 0, \tag{18}$$

or

$$\zeta + 2g\xi + 2f\eta + c(1-\zeta) = 0, \tag{19}$$

or

$$(1-c)\zeta + 2g\xi + 2f\eta + c = 0. \tag{20}$$

This equation is that of a plane, and the curve of intersection of this plane and the given sphere is a circle. We may therefore conclude that by stereographic projection circles in the complex plane become circles upon the sphere. These circles do not in general pass through the origin nor through the north pole of the sphere, as we may see from an examination of equation (20).

A general property of stereographic projection is stated in the following theorem.

THEOREM. *The mapping of the sphere upon the complex plane, and conversely, by means of stereographic projection is conformal.*

It has been pointed out that the general condition for conformal mapping is that we have

$$ds = M \cdot dS,$$

where ds, dS are differential elements of arcs upon the two surfaces concerned. From the calculus of real variables we have

$$dS^2 = d\xi^2 + d\eta^2 + d\zeta^2, \tag{21}$$

where dS relates to the sphere, and moreover we have

$$ds^2 = dx^2 + dy^2, \tag{22}$$

where ds is taken in the complex plane.

From (13) we obtain

$$dx = \frac{d\xi}{1-\zeta} + \frac{\xi\, d\zeta}{(1-\zeta)^2},$$

$$dy = \frac{d\eta}{1-\zeta} + \frac{\eta\, d\zeta}{(1-\zeta)^2}.$$

Hence

$$ds^2 = \frac{d\xi^2 + d\eta^2}{(1-\zeta)^2} + 2\,\frac{(\xi\, d\xi + \eta\, d\eta)\, d\zeta}{(1-\zeta)^3} + \frac{(\xi^2+\eta^2)\, d\zeta^2}{(1-\zeta)^4}. \tag{23}$$

From equation (2) we get

$$\xi^2 + \eta^2 = \zeta(1-\zeta), \tag{24}$$

whence

$$2(\xi\,d\xi + \eta\,d\eta) = d\zeta - 2\,\zeta\,d\zeta. \tag{25}$$

Substituting these values in (23), we obtain

$$ds^2 = \frac{d\xi^2 + d\eta^2 + d\zeta^2}{(1-\zeta)^2},$$

from which we get

$$ds = \frac{1}{1-\zeta}\cdot dS.$$

The definition of conformal mapping is therefore satisfied, and the ratio of magnification M in passing from the sphere to the complex plane is in this case $\frac{1}{1-\zeta}\cdot$ Similarly, it may be shown that

$$dS = \frac{1}{x^2+y^2+1}\,ds,$$

and hence the mapping from the complex plane upon the sphere is also conformal, having $\frac{1}{x^2+y^2+1}$ as the ratio of magnification. As we might expect, this ratio of magnification becomes infinite at the point $\zeta = 1$, that is at the north pole.

41. Classification of linear fractional transformations. The geometrical interpretation of the linear fractional transformations of the complex plane into itself may be regarded as a problem in kinematics. In the present article we shall undertake to classify these transformations of the plane by means of the corresponding motions of the points of the plane. We have already seen that the linear transformation given by an equation of the form

$$\bar{w} = \bar{z} + \mu \tag{1}$$

is a translation of the points of the complex plane. Suppose the lines of motion be the system of parallel lines AB, Fig. 77. Let this system of straight lines be mapped by reciprocation with respect to the origin. As such a reciprocation consists of geometric inversion with respect to a unit circle about the origin followed by reflection upon the axis of reals, the resulting configuration is a system of coaxial circles through the origin having a common tangent at that point. The particular line $\bar{L}_1$ of the system AB of straight lines which passes through the origin maps into that straight line L_1 through the origin which is the reflection of the given line with respect to the axis of

reals. Those lines lying below $\bar{L}_1$ map into circles tangent to L_1 at the origin and lying above it. Likewise the lines of AB lying above $\bar{L}_1$ map into circles tangent to L_1 at the origin and lying below L_1. Corresponding to a motion along the lines AB, we have a motion of the points along this system of circles through the origin. The corresponding directions of the motions in the two cases are indicated by the arrow-heads. The orthogonal lines CD map by the same reciprocation into a system of coaxial circles through the origin and orthogonal to the first system of circles as indicated in Fig. 77. The

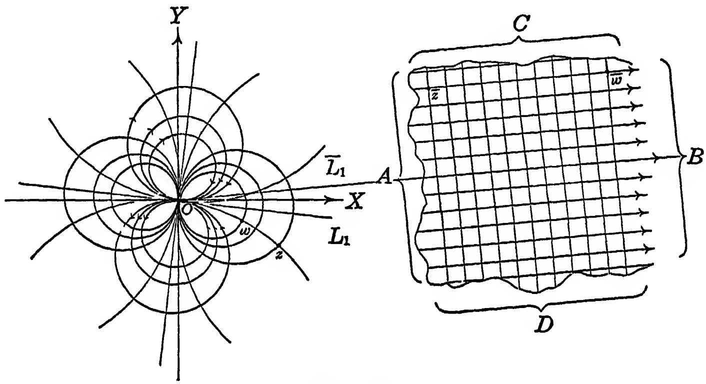

FIG. 77.

motion of the plane as here indicated is called a **parabolic motion** about the origin.

That a parabolic motion about the origin is a linear fractional transformation of the plane may be easily shown. Let z and w be the initial and final positions, respectively, of a point moving along one of these circles. Corresponding to these two points, we have two points $\bar{z}$, $\bar{w}$ upon some line of the system AB, given by the relations

$$\bar{z} = \frac{1}{z}, \qquad \bar{w} = \frac{1}{w}. \tag{2}$$

As we have seen, the points $\bar{z}$, $\bar{w}$ are associated by the relation given in (1). Substituting in this equation the values of $\bar{z}$, $\bar{w}$ given in (2), we have

$$\frac{1}{w} = \frac{1}{z} + \mu,$$

or
$$w = \frac{z}{\mu z + 1}, \tag{3}$$

which is a linear fractional relation between w and z.

The points remaining invariant by a parabolic motion about the origin may be readily found; for, by comparing (3) with the general form of the linear fractional transformation, we have

$$\alpha = 1, \qquad \beta = 0, \qquad \gamma = \mu, \qquad \delta = 1.$$

Hence, from Art. 39, we have as the invariant points

$$z_1 = z_2 = 0;$$

that is, the two invariant points are coincident at the origin.

If the parabolic motion takes place about any point $z_0 \neq 0$, the relation between the initial and final values of the variable, namely between z and w, is still linear. For the translation

$$z_1 = z - z_0, \qquad w_1 = w - z_0 \tag{4}$$

brings the origin to the point z_0, and z_1, w_1 are respectively the initial and final values represented by the variable point with respect to z_0. Since the motion about z_0 is parabolic, we have

$$w_1 = \frac{z_1}{1 + \mu z_1}.$$

Putting for z_1, w_1 their values in (4) we have

$$w - z_0 = \frac{z - z_0}{1 + \mu(z - z_0)},$$

or
$$w = \frac{z(1 + \mu z_0) - \mu z_0^2}{\mu z + (1 - \mu z_0)},$$

which is a linear fractional relation. The points left invariant by a parabolic motion about z_0 are coincident at z_0.

In case of a translation, the invariant points are given by (1) and are coincident at infinity. Consequently, a translation may be regarded as a special case of a parabolic motion where the invariant points coincide at infinity. But as we have seen, a translation is a linear fractional transformation. We may then conclude that every parabolic motion corresponds to a linear fractional transformation having coincident invariant points.

We shall now show that conversely every linear transformation having two coincident invariant points is a parabolic motion. First

of all, suppose the two invariant points coincide at infinity. Then we must have

$$(\alpha - \delta)^2 + 4\beta\gamma = 0, \qquad \gamma = 0,$$

whence,

$$\alpha = \delta.$$

The general linear fractional transformation then reduces to

$$w = z + \frac{\beta}{\alpha},$$

which is as we know a translation, that is a special case of a parabolic motion.

If the two invariant points coincide at a finite point z_0, then by translation the origin can be moved to this point. But a translation does not change the form of the lines of motion. By reciprocation these lines of motion through the origin map into lines of motion having two coincident invariant points at $z = \infty$. But as we have seen such a motion is a translation, and by definition the motion along the reciprocal of these lines is a parabolic motion about the origin and consequently the original motion is a parabolic motion about the point z_0.

Another important class of motions is obtained when we apply the reciprocal substitution to the general linear transformation, which may be written

$$\overline{w} = \nu\overline{z} + \mu, \tag{5}$$

where $\overline{z}$, $\overline{w}$ are respectively the initial and final positions of the variable z-point, and ν, μ are complex constants.

By the reciprocal substitution, the point P maps into some point P', Fig. 78. The pencil of rays passing through P maps into circles through P' and through the inverse of the point at infinity, namely the origin O. If the transformation is such that the half-rays are the lines of motion in the one case, then the corresponding lines of motion in the other case are the circles passing through P' and O. The direction of the motion is indicated by the arrow-heads. The resulting motion is called a **hyperbolic motion** through the fixed points O and P'.

If the concentric circles about P are considered as the lines of motion, then in the reciprocal configuration the circles having their centers on OP' are lines of motion and the circles passing through O

and P' form the orthogonal system. The motion in this case is called an **elliptic motion.**

By a combination of rotation and stretching, we have, as we have seen, a logarithmic spiral motion. Corresponding to the logarithmic

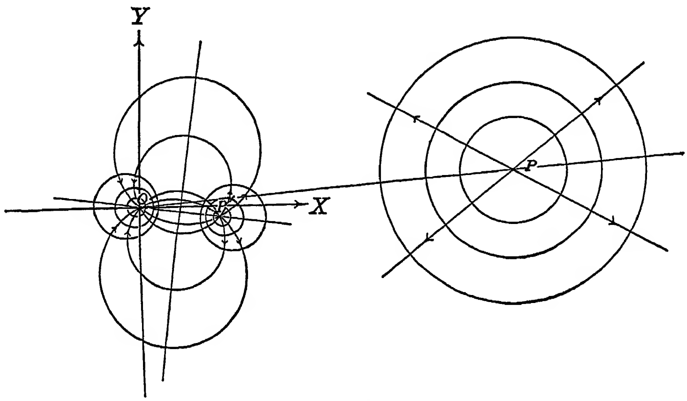

FIG. 78.

motion about P we have after reciprocation with respect to the origin what we shall call a **loxodromic motion** about the points P' and O as indicated in Fig. 79.

Since hyperbolic and elliptic motions appear as special cases of a loxodromic motion, it follows that all three motions are linear

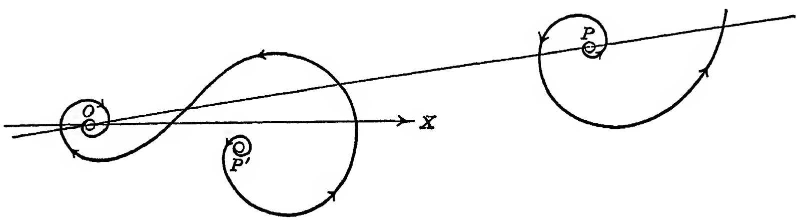

FIG. 79.

fractional transformations if it can be shown that a loxodromic motion is such a transformation. To show this, let as before z, w be the initial and the final positions, respectively, of the variable point on one of the curves of a loxodromic motion about O and P'. We have then

$$\bar{z} = \frac{1}{z}, \qquad \bar{w} = \frac{1}{w}, \tag{6}$$

where $\bar{z}$, $\bar{w}$ are the points on the logarithmic spiral corresponding to z and w. The points $\bar{z}$, $\bar{w}$ are related to each other by equation (5). Substituting for $\bar{z}$, $\bar{w}$ their values, we get

$$\frac{1}{w} = \frac{\nu}{z} + \mu,$$

or

$$w = \frac{z}{\mu z + \nu}, \tag{7}$$

which is the linear fractional relation required.

We have, by comparison with the general linear fractional transformation,

$$\alpha = 1, \qquad \beta = 0, \qquad \gamma = \mu, \qquad \delta = \nu.$$

Hence, if $\nu \neq 1$, we get

$$(\alpha - \delta)^2 + 4\beta\gamma \neq 0, \qquad \gamma \neq 0;$$

that is the invariant points are finite and distinct. From equation (3), Art. 39, it will be seen that the two invariant points are 0 and $\frac{1-\nu}{\mu}$. Instead of the origin being taken as one of the invariant points, any point P_1 may be selected. The second invariant point is then obtained by reciprocation of the point P with respect to a unit circle about the point P_1 as a center. But to remove the origin to a given point is a translation of the points of the plane, which as we know is a linear substitution. Hence it can be shown by the same method as that employed in the discussion of parabolic motion that a loxodromic motion about any two fixed points of the plane is a linear fractional transformation having those points as the invariant points. If one of the invariant points is $z = \infty$, we have then $\gamma = 0$ and the transformation reduces to

$$w = \frac{\alpha z + \beta}{\delta} = \frac{\alpha}{\delta} z + \frac{\beta}{\delta},$$

which gives, as we have seen, a logarithmic spiral motion. We can then regard a logarithmic spiral motion as a special case of a loxodromic motion, where one of the distinct invariant points is at infinity.

Conversely, every linear fractional transformation of the points of the complex plane such that two distinct points are left invariant is some form of a loxodromic motion. If one of these points is finite while the other is at infinity we have a logarithmic spiral motion,

which is as we have seen a special kind of loxodromic motion. If both invariant points are finite, then by a translation one of these points can be made the origin and by reciprocation with respect to this new origin the original lines of motion are mapped into lines of motion having two distinct invariant points, one being at infinity, that is into a logarithmic spiral motion. Hence, by the definition of a loxodromic motion the original motion is of that type, since the translation which moves one of the invariant points to the origin does not affect the form of the lines of motion.

It has already been pointed out that a hyperbolic motion and an elliptic motion are special cases of a loxodromic motion. The conditions under which these special cases arise depend upon the character of ν, as may be seen from an examination of (5). Suppose we put

$$\nu = \rho(\cos\phi + i \sin\phi).$$

Then, if $\rho \neq 0$, $\phi \neq 0$, the logarithmic spiral motion maps into the loxodromic motion about the finite invariant points; if $\rho \neq 1$, $\phi = 0$, we have from (5) a pencil of rays and after mapping we get a hyperbolic motion; if $\rho = 1$, $\phi \neq 0$, the resulting motion is elliptic. It will also be observed, that in case $\nu = 1$, the motion represented by (5) is a translation, in which case the invariant points become coincident at infinity and after mapping by reciprocation with respect to some finite point the resulting motion is the parabolic motion already discussed.

We are now in a position to classify the various types of linear fractional transformations of a plane according to their kinematic properties. For, as a result of the foregoing discussion, it follows that every linear fractional transformation may be interpreted in terms of some one of the motions already discussed. As we have seen, every linear fractional transformation leaves two points of the plane invariant. These points must be either coincident, or distinct. If they coincide, we have seen that the transformation is a parabolic motion, where a translation appears as a special case. If the invariant points are finite and distinct, we have in general a loxodromic motion, which reduces to a hyperbolic motion or an elliptic motion according to the particular value taken by ν. If the invariant points are distinct, one being at infinity, then our transformation is a logarithmic spiral motion, which again reduces to an expansion or a rotation according to the particular values given to ν as already noted.

EXERCISES

1. Given in the Z-plane the two intersecting curves $x^2 + (y - 2)^2 = 4$ and $2x + 3y - 7 = 0$. Map these curves upon the W-plane by means of the relation $w = \frac{1}{z}$. Show that the resulting curves intersect at the same angle as the given curves.

2. Map the orthogonal systems of straight lines $u = c$, $v = c$ from the W-plane upon the Z-plane; first, by means of the relation $w = z^2$, second by means of the relation $w = \frac{1}{z^2}$. How would the two resulting configurations be related if projected stereographically?

3. Given the ellipse $\frac{x^2}{4} + \frac{y^2}{9} = 1$. Subject this curve to the transformation $w = 3z + 10$ and find the equation of the resulting curve. Draw the curve. What general principle do the results of this mapping illustrate?

4. By applying the Cauchy-Riemann differential equations to the relation $w = \frac{1}{z}$, show that w is an analytic function

5. A point z moves with uniform speed of 2 cm. per sec. along a circle whose center is at $2 + 3i$ and whose radius is 4 cm. Determine the path, velocity, and acceleration of w, where $w = 2z + (1 + 2i)$.

6. Given the three points $1 + 2i$, $2 + i$, $2 + 3i$. Determine the linear fractional transformation that will map these points into the following points: $2 + 2i$, $1 + 3i$, 4. Find the invariant points of the transformation.

7. Let a circle passing through the three points $2 + 3i$, 0, $3 + 2i$ be given. Find the conjugate of the point $1 + 2i$ with respect to this circle. Construct the curve into which the circle through the points $1 + 2i$, $2 + 3i$, $3 + 2i$ is changed by reciprocation with respect to the circle of radius 2 about the origin.

8. A given translation of the complex plane is represented by the equation $w = z + \beta$, where $\beta = 2 + 3i$. Show in two ways how the lines of motion in the complex plane can be changed into lines of parabolic motion upon the sphere having the coincident invariant points at the north pole. Explain why the two methods give the same result.

9. The speed at the point $2 + 4i$ of a moving point in a motion of expansion about the point $1 + 2i$ is 2 units per sec. At the same instant what is the position and velocity of the corresponding point in the hyperbolic motion obtained through reciprocation of this motion of expansion?

10. When a transformation has the property that on being repeated once it restores every point of the plane to its initial position, it is called an **involution.** Determine the conditions that must exist between the coefficients α, β, γ, δ, in order that the transformation

$$w = \frac{\alpha z + \beta}{\gamma z + \delta}$$

shall be an involution.

11. Show that every involution is an elliptic transformation.

12. Show how the regions for the anharmonic ratios of four points may be represented on the complex plane in case the four points are harmonic; in case they are equianharmonic.

CHAPTER VI

INFINITE SERIES

42. Series of complex terms. In this chapter we shall consider some of the general properties of infinite series whose terms are complex numbers. We shall assume such knowledge of infinite series as is usually given in elementary texts on algebra and calculus. Special attention will be directed to the properties of power series.

Suppose we have given the series

$$\Sigma\alpha_n = \alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_n + \cdots, \tag{1}$$

where $\alpha_n = a_n + ib_n$, a_n, b_n, being real numbers. As in series of real terms, we shall denote the sum of the first n terms by S_n, that is, we shall put

$$S_n = \alpha_1 + \alpha_2 + \cdots + \alpha_n.$$

A series is said to be **convergent** if the sequence

$$S_1, S_2, \ldots, S_n, \ldots$$

has a limit. If α is this limit, that is if

$$\underset{n=\infty}{L} S_n = \alpha,$$

then the series is said to converge to the number α. This number is also called the **sum** of the series and is uniquely determined by the series. We may therefore write

$$\Sigma\alpha_n = \alpha.$$

If the limit of S_n does not exist as n increases indefinitely, then we say that the given series is **divergent.**

The geometric significance of convergence may be easily seen. All that is needed is to locate in the complex plane the points that represent

$$\begin{aligned}
S_1 &= \alpha_1,\\
S_2 &= \alpha_1 + \alpha_2,\\
S_3 &= \alpha_1 + \alpha_2 + \alpha_3,\\
&\cdots\cdots\cdots\\
S_n &= \alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_n,\\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

It will be noticed that these points do not necessarily lie along a straight line as in series of real terms, but that they may be distributed over the complex plane. They are, however, so located as to be dense at the limiting point α.

In order that the limit shall exist, it is necessary that $\underset{n=\infty}{L} |\alpha_n| = 0.$ As in series of real terms, this condition is a necessary but not a sufficient condition for the convergence of the given series.

As already stated, each term of a series of complex numbers may be written in the form

$$\alpha_n = a_n + ib_n.$$

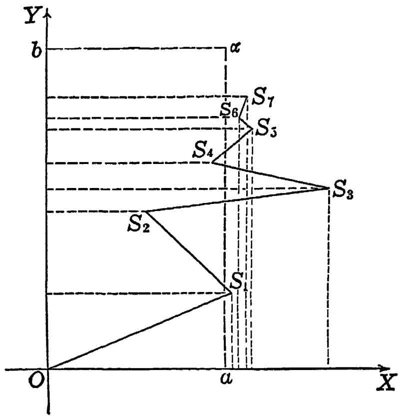

FIG. 80.

Separating the real and the imaginary parts of the terms of the series, we may put

$$A_n = a_1 + a_2 + \cdots + a_n,$$
$$B_n = b_1 + b_2 + \cdots + b_n.$$

We may now formulate the following theorem.

THEOREM I. *The necessary and sufficient condition that the series of complex numbers*

$$\alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_n + \cdots$$

converges to a limit $\alpha = a + ib$ *is that*

$$\underset{n=\infty}{L} A_n = a, \qquad \underset{n=\infty}{L} B_n = b. \tag{2}$$

We have

$$S_n = A_n + iB_n,$$

and by Theorem I, Art. 12, the necessary and sufficient condition that the sequence

$$S_1, S_2, S_3, \ldots, S_n, \ldots$$

has a limit $\alpha = a + ib$ is that

$$\underset{n=\infty}{L} A_n = a, \qquad \underset{n=\infty}{L} B_n = b.$$

Hence the theorem.

The point a is determined upon the axis of reals (Fig. 80) by the series Σa_n, while b is determined upon the axis of imaginaries by the series Σb_n. The limiting point α is then identical with $a + ib$.

It will be seen at once, therefore, that if either of the series Σa_n, Σb_n is divergent, the series

$$\Sigma \alpha_n \equiv \Sigma(a_n + ib_n)$$

is divergent.

Ex. 1. Given the series $\sum\left(\frac{1}{n}+\frac{i}{2^n}\right)=\left(1+\frac{1}{2}i\right)+\left(\frac{1}{2}+\frac{1}{4}i\right)+\left(\frac{1}{3}+\frac{1}{8}i\right)+\cdots.$

This series is divergent; for, we have

$$\sum a_n = 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots,$$

$$\sum b_n = \frac{1}{2}+\frac{1}{2^2}+\frac{1}{2^3}+\frac{1}{2^4}+\cdots,$$

where Σa_n is divergent although Σb_n is convergent.

We may also state the necessary and sufficient condition for the convergence of a series of complex terms as follows:

THEOREM II. *The necessary and sufficient condition that the series*

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n + \cdots$$

converges is that for an arbitrarily small positive number ϵ there exists a definite number m such that

$$|\alpha_{n+1} + \alpha_{n+2} \cdots + \alpha_{n+p}| < \epsilon, \quad n > m, \quad p = 1, 2, 3, \ldots.$$

This condition follows at once from Theorem VI, Art. 12. We have the sequence

$$S_1, S_2, S_3, \ldots, S_n, \ldots.$$

The necessary and sufficient condition for the convergence of this sequence is that, for an arbitrarily small ϵ, a corresponding integer m may be found such that

$$|S_{n+p} - S_n| < \epsilon, \quad n > m, \quad p = 1, 2, 3, \ldots.$$

Replacing the values of S_n and S_{n+p} by their values in terms of the α's, we have at once the condition stated in the theorem.

THEOREM III. *The series $\Sigma\alpha_n$ converges, provided the series of moduli of the various terms of the given series converges.*

We have by hypothesis the condition that the series of absolute values

$$|\alpha_1| + |\alpha_2| + \cdots + |\alpha_n| + \cdots$$

converges. From Theorem II it follows that since the foregoing series of moduli converges, we have

$$\{\, |\alpha_{n+1}| + \cdots + |\alpha_{n+p}| \,\} < \epsilon, \qquad n > m, \qquad p = 1, 2, 3, \cdots.$$

But we have

$$|\alpha_{n+1} + \cdots + \alpha_{n+p}| \leqq |\alpha_{n+1}| + \cdots + |\alpha_{n+p}|,$$

whence it follows that

$$|\alpha_{n+1} + \cdots + \alpha_{n+p}| < \epsilon, \qquad n > m, \qquad p = 1, 2, 3, \cdots.$$

By Theorem II this inequality gives a sufficient condition for the convergence of the given series $\Sigma\alpha_n$.

Ex. 2. Let it be required to test the convergence of the series $\Sigma\alpha_n$, where $\alpha_n = \dfrac{\rho^n(\cos n\theta + i \sin n\theta)}{n}$.

The series of moduli is

$$\rho + \frac{\rho^2}{2} + \frac{\rho^3}{3} + \cdots + \frac{\rho^n}{n} + \cdots,$$

which converges for $\rho < 1$; for, we have by the ratio test

$$\underset{n=\infty}{L} \left\{ \frac{\rho^{n+1}}{n+1} : \frac{\rho^n}{n} \right\} = \underset{n=\infty}{L}\, \rho \cdot \frac{n}{n+1} = \rho < 1.$$

Hence, by Theorem III the given series converges for $|\alpha_n| < 1$.

If the series of absolute values of the various terms of the series $\Sigma\alpha_n$, namely $\Sigma\,|\alpha_n| = \Sigma\sqrt{a_n^2 + b_n^2}$, converges, then the given series $\Sigma\alpha_n$ is said to **converge absolutely.**

THEOREM IV. *Given the series $\Sigma\alpha_n$, where $\alpha_n = a_n + ib_n$. The necessary and sufficient condition that this series converges absolutely is that the two series Σa_n, Σb_n converge absolutely.*

That the absolute convergence of Σa_n and Σb_n is a necessary condition may be shown as follows. We have

$$|a_n| \leqq |a_n + ib_n| = \sqrt{a_n^2 + b_n^2}, \quad |b_n| \leqq |a_n + ib_n| = \sqrt{a_n^2 + b_n^2}. \quad (3)$$

We assume that the given series converges absolutely; that is, that the series of moduli converges. We have then the convergent series

$$\Sigma\,|\alpha_n| = \sqrt{a_1^2 + b_1^2} + \cdots + \sqrt{a_n^2 + b_n^2} + \cdots. \qquad (4)$$

From (3) it follows at once that the terms of the two series $\Sigma\mid a_n\mid$, $\Sigma\mid b_n\mid$ are equal to or less than the corresponding terms of the convergent series (4) and hence both of the series Σa_n, Σb_n converge absolutely.

The condition set forth in the theorem is also sufficient. If we write

$$A'_n = |a_1| + |a_2| + \cdots + |a_n|,$$
$$B'_n = |b_1| + |b_2| + \cdots + |b_n|,$$

we have

$$\underset{n=\infty}{L}(A'_n + B'_n) = \underset{n=\infty}{L} A'_n + \underset{n=\infty}{L} B'_n;$$

that is, since $\Sigma\mid a_n\mid$, $\Sigma\mid b_n\mid$ converge, the series $\Sigma\,\{\mid a_n\mid + \mid b_n\mid\}$ also converges. We have, however,

$$|\alpha_n| \leqq |a_n| + |b_n|,$$

and consequently the series $\Sigma\mid\alpha_n\mid$ converges; that is, the given series converges absolutely.

Ex. 3. Test the series $\sum \frac{i^n}{n}$ for absolute convergence.

We may write the given series in the form

$$i + \frac{i^2}{2} + \frac{i^3}{3} + \cdots = i - \frac{1}{2} - \frac{i}{3} + \frac{1}{4} + \frac{i}{5} - \frac{1}{6} - \cdots,$$

from which we have

$$\sum a_n = -\left(\frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \cdots\right),$$
$$\sum b_n = \left(1 - \frac{1}{3} + \frac{1}{5} - \cdots\right).$$

Each of these series is convergent, but neither converges absolutely. Hence the given series can not converge absolutely. In fact we have

$$\sum \left|\frac{i^n}{n}\right| = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} + \cdots,$$

which is a well-known divergent series.

While absolute convergence, as we have seen, gives a sufficient condition for the convergence of $\Sigma\alpha_n$, it is not a necessary condition. It may occur, as we have just seen (Ex. 3), that a series converges even if the series of moduli is divergent.

Ex. 4. Given the series $\sum \frac{z^n}{n}$, where $z = \cos\theta + i\sin\theta$, $0 < \theta < 2\pi$.

It will be observed that the series in Ex. 3 is a special case of the series here

considered, namely, where $\theta = \frac{\pi}{2}$. The series of moduli is $\sum \frac{1}{n}$, which is a divergent series. However, the given series

$$\Sigma \alpha_n = \Sigma(a_n + ib_n)$$

is convergent, since each of the two series

$$\sum a_n = \sum \frac{1}{n} \cos n\theta = \cos\theta + \frac{\cos 2\theta}{2} + \frac{\cos 3\theta}{3} + \cdots,$$

$$\sum b_n = \sum \frac{1}{n} \sin n\theta = \sin\theta + \frac{\sin 2\theta}{2} + \frac{\sin 3\theta}{3} + \cdots$$

is convergent for $0 < \theta < 2\pi$.*

Absolutely convergent series have certain properties not possessed by series in general. Some of these properties are stated in the following theorems.

THEOREM V. *The sum of an absolutely convergent series of complex terms is independent of the order of the terms.*

When we have a finite series the order of arrangement of the terms is a matter of indifference. The foregoing theorem enables us to extend this commutative property to an infinite series, provided it converges absolutely.

Let the given series be

$$\Sigma \alpha_n = \alpha_1 + \alpha_2 + \cdots + \alpha_n + \cdots, \tag{5}$$

where $\alpha_n = a_n + ib_n$.

By Theorem IV the two series Σa_n, Σb_n converge absolutely. When a series of real terms converges absolutely the sum of the series is not dependent upon the order in which the terms of the series are arranged.† Since we have

$$\Sigma \alpha_n = \Sigma(a_n + ib_n) = \Sigma a_n + i\Sigma b_n \tag{6}$$

and the terms of the two series Σa_n, Σb_n can be taken in any order without affecting their sum, it follows from (6) that the sum of the given series $\Sigma \alpha_n$ is likewise independent of the order in which its terms are taken.

The associative law of addition may be extended to any convergent infinite series; that is, we can always insert parentheses at pleasure in an infinite series. However, if the series converges absolutely the

* See Bromwich, *Theory of Infinite Series*, p. 159.

† *Ibid.*, p. 64.

sum is not affected by any arbitrary rearrangement and grouping of the terms; that is, we have the following theorem.

THEOREM VI. *The sum of an absolutely convergent series remains unchanged if the terms are rearranged and grouped in any arbitrary manner.*

Let $\Sigma\alpha_n$ be any absolutely convergent series having the sum α. Let us first suppose the terms of this series to be grouped by putting into each group a certain number of consecutive terms. Let A_1, A_2, . . . denote these groups, respectively, where

$$\begin{aligned} A_1 &= \alpha_1 + \cdots + \alpha_k, \\ A_2 &= \alpha_{k+1} + \cdots + \alpha_r, \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

We shall now establish the convergence of the series

$$A_1 + A_2 + \cdots + A_n + \cdots. \tag{7}$$

The sum of the first n terms of this series, namely,

$$S_n' = A_1 + \cdots + A_n,$$

is equal to the sum of the first m terms $(m > n)$ of the given series $\Sigma\alpha_n$. As n becomes infinite, m also increases without limit. We have for all values of n

$$S_n' = S_m, \tag{8}$$

where S_m denotes the sum of the first m terms of the given series $\Sigma\alpha_n$. However, m does not take all integral values, but increases through the values k, r, s, ... Since the given series converges, it follows by Art. 12 that the limiting value of S_m is the sum of the given series, namely α. As (8) holds for all values of n, it follows that we have also

$$\underset{n=\infty}{L}\, S_n' = \alpha.$$

Since the given series converges absolutely the sum of the series is not affected by any arbitrary rearrangement of its terms. Consequently, in order to get any desired grouping of the terms of the given series $\Sigma\alpha_n$, all that is necessary is to so rearrange the terms of the series that the terms occur in the required order and then form the series (7) of groups of consecutive terms. Hence the theorem.

Convergent series that do not converge absolutely are called **conditionally convergent** series. The following theorem may be stated with reference to such series, namely:

THEOREM VII. *The sum of a conditionally convergent series depends upon the order of arrangement of its terms.*

Let $\Sigma\alpha_n$ be the given conditionally convergent series. As in the demonstration of Theorem V, we have

$$\Sigma\alpha_n = \Sigma(a_n + ib_n) = \Sigma a_n + i\Sigma b_n.$$

By Theorem IV we know that at least one of the series Σa_n, Σb_n must converge conditionally; otherwise the given series would converge absolutely. Suppose that Σb_n alone converges conditionally. It follows from (6) that the given series $\Sigma\alpha_n$ converges to a limit $a + ix$, where $x = \Sigma b_n$ depends upon the arrangement of the terms of the series Σb_n, and by properly choosing this arrangement x may be made any arbitrarily chosen real number.* This establishes the theorem.

In this demonstration we have made use of the fact that in a conditionally convergent series of real terms the series may be made to converge to any arbitrarily chosen real number. This fact would seem to suggest that a conditionally convergent series of complex terms might likewise be made to converge to any arbitrarily chosen complex number as a limit. This, however, is not in general the case. As already pointed out, if Σb_n converges conditionally, we may so arrange the terms of this series as to make it converge to any arbitrarily chosen number x. In order that the series $\Sigma\alpha_n$, by a rearrangement of its terms, shall approach an arbitrarily chosen complex number $x + iy$ it is, in general, also necessary that Σa_n be conditionally convergent and that the rearrangement of its terms be unrestricted. This, however, is impossible, for when the b's are properly arranged the a's must be taken in such an order that there is associated with each b_k an a_k such that $a_k + ib_k$ is one of the α's. Since the choice of the arrangement of the a's is thus restricted, it follows that while the limit of a conditionally convergent series of complex terms depends upon the order of its terms, it can not, in general, be made to approach any complex number at pleasure by a suitable arrangement of its terms. Some particular conditionally convergent series of complex terms may, however, be made to approach any previously assigned limit by the proper arrangement of the terms.†

* See Bromwich, *Theory of Infinite Series*, p. 68.

† See Lévy, *Nouv. Ann. Math.*, Vol. 5, 1905, p. 506.

To test the convergence of a series of complex terms, we need only to employ the methods of testing series of real terms. We can always make the convergence of any series $\Sigma\alpha_n$ depend upon the convergence of the two series of real terms Σa_n, Σb_n, where $\alpha_n = a_n + ib_n$, by use of the Theorem I. Frequently, it is more convenient to test the convergence of the given series by considering the series of moduli. As we have seen, if this series converges then the given series converges. The series of moduli is a series of real positive terms and the tests for the convergence of such series may be at once applied. If the series of complex terms converges conditionally, we have no general tests; but this is to be expected, since no general tests exist for the conditional convergence of series whose terms are real.

In the discussion thus far we have considered only those series whose terms are constants. The terms of the series may, however, be functions of a complex variable. Such a series may be written

$$u_1(z) + u_2(z) + \cdots + u_n(z) + \cdots.$$

The convergence of this series for a given value of the variable implies that if this value is substituted for the variable, the resulting series of constants converges. The region of convergence may or may not be a closed region. A series of functions of a complex variable defines a function of the complex variable in the region of convergence.

43. Operations with series. In order to make use of series in mathematical discussions, it is necessary to establish the conditions under which convergent infinite series may be combined by the fundamental operations of arithmetic following the formal laws applicable to sums of a finite number of terms. It follows at once from the laws governing operations with limits and from the definition of convergence, that any convergent series may be multiplied or divided termwise by a constant, and that a constant may be added to or subtracted from any convergent infinite series by adding it to or subtracting it from any term of this series, precisely as in the case of a sum of a finite number of terms. We shall now consider the sum, difference, product, and quotient of two convergent infinite series.

THEOREM I. *If $\Sigma\alpha_n$, $\Sigma\beta_n$ are two convergent series having the limits α, β, respectively, then*

$$\alpha+\beta=\Sigma(\alpha_n+\beta_n)=(\alpha_1+\beta_1)+(\alpha_2+\beta_2)+\cdots+(\alpha_n+\beta_n)+\cdots, \quad (1)$$

If the given series converge absolutely, then the series (1) *converges absolutely.*

Let
$$S_n = \alpha_1 + \alpha_2 + \cdots + \alpha_n,$$
$$T_n = \beta_1 + \beta_2 + \cdots + \beta_n.$$
Then, we have
$$S_n + T_n = (\alpha_1 + \beta_1) + (\alpha_2 + \beta_2) + \cdots + (\alpha_n + \beta_n). \tag{2}$$
By hypothesis we have
$$\underset{n=\infty}{L} S_n = \alpha, \qquad \underset{n=\infty}{L} T_n = \beta;$$
hence, it follows that
$$\underset{n=\infty}{L} (S_n + T_n) = \underset{n=\infty}{L} S_n + \underset{n=\infty}{L} T_n = \alpha + \beta,$$
whence
$$\alpha + \beta = (\alpha_1 + \beta_1) + (\alpha_2 + \beta_2) + \cdots + (\alpha_n + \beta_n) + \cdots.$$

We shall now show that if $\Sigma\alpha_n$, $\Sigma\beta_n$ converge absolutely, the series (1) converges absolutely. Let
$$|\alpha_n| = \rho_n, \quad |\beta_n| = r_n, \quad \Sigma\rho_n = \rho, \quad \Sigma r_n = r,$$
$$S_n' = \rho_1 + \rho_2 + \cdots + \rho_n,$$
$$T_n' = r_1 + r_2 + \cdots + r_n.$$
Putting
$$p_n = \rho_n + r_n,$$
we may write
$$P_n = p_1 + p_2 + \cdots + p_n.$$
We have then
$$\underset{n=\infty}{L} P_n = \underset{n=\infty}{L} (S_n' + T_n') = \rho + r;$$
that is,
$$\rho + r = (\rho_1 + r_1) + (\rho_2 + r_2) + \cdots + (\rho_n + r_n) + \cdots. \tag{3}$$
Since we have
$$|\alpha_n + \beta_n| \leqq |\alpha_n| + |\beta_n| = \rho_n + r_n,$$
it follows from (3) that the series of moduli of (2) converges, that is (2) converges absolutely.

It is to be noted that the parentheses in (1) may be removed; for the sum of the first n terms of the series
$$\alpha_1 + \beta_1 + \alpha_2 + \beta_2 + \alpha_3 + \cdots$$
differs from the sum of a properly chosen number of terms of (1) by at most the first term in one of the parentheses, this term approaching zero with increasing n.

Ex. 1. Add the two series

$$\log(1+z) = z - \tfrac{1}{2}z^2 + \tfrac{1}{3}z^3 - \tfrac{1}{4}z^4 + \tfrac{1}{5}z^5 - \cdots,$$
$$\log(1-z) = -z - \tfrac{1}{2}z^2 - \tfrac{1}{3}z^3 - \tfrac{1}{4}z^4 - \tfrac{1}{5}z^5 - \cdots.$$

Applying the results of Theorem I, we obtain

$$\begin{aligned}\log(1-z^2) &= \log(1+z) + \log(1-z)\\ &= -z^2 - \tfrac{1}{2}z^4 - \tfrac{1}{3}z^6 - \cdots.\end{aligned}$$

The corresponding theorem for the difference of two series may be stated as follows:

THEOREM II. *If $\Sigma\alpha_n$, $\Sigma\beta_n$ are two convergent series having the limits α and β respectively, then*

$$\alpha - \beta = \Sigma(\alpha_n - \beta_n) = (\alpha_1 - \beta_1) + (\alpha_2 - \beta_2) + \cdots + (\alpha_n - \beta_n) + \cdots. \quad (4)$$

If the given series converge absolutely, then the series (4) *converges absolutely.*

The first part of this theorem may be established by a method similar to that employed in the demonstration of Theorem I and the demonstration need not be repeated. To show that series (4) converges absolutely, we have

$$|\alpha_n - \beta_n| \leqq |\alpha_n| + |\beta_n| = \rho_n + r_n. \quad (5)$$

In the demonstration of Theorem I, it was shown that the series $\Sigma(\rho_n + r_n)$ converges. Hence from (5) it follows that the series $\Sigma\,|\alpha_n - \beta_n|$ must converge; that is (4) converges absolutely.

Ex. 2. From the series

$$\log(1+z) = z - \tfrac{1}{2}z^2 + \tfrac{1}{3}z^3 - \tfrac{1}{4}z^4 + \tfrac{1}{5}z^5 - \cdots$$

subtract the series

$$\log(1-z) = -z - \tfrac{1}{2}z^2 - \tfrac{1}{3}z^3 - \tfrac{1}{4}z^4 - \tfrac{1}{5}z^5 - \cdots.$$

By Theorem II we get

$$\log\left(\frac{1+z}{1-z}\right) = \log(1+z) - \log(1-z) = 2z + \frac{2}{3}z^3 + \frac{2}{5}z^5 + \cdots.$$

The following theorem due to Cauchy * gives a convenient criterion for the multiplication of series.

THEOREM III. *If the two series $\Sigma\alpha_n$, $\Sigma\beta_n$ converge absolutely to the limits α and β, respectively, and if each term of the one series is multiplied into each term of the other, the series whose terms are these products taken in any order which includes them all in a simple infinite series converges absolutely and has the limit $\alpha\beta$.*

* *Enc. der Math. Wiss.*, IA_3, p. 96; *Enc. des Sci. Mat.*, I_4, p. 247.

If each term of $\Sigma\alpha_n$ is multiplied into each term of $\Sigma\beta_n$ the result may be schematically displayed as follows:

$$\left.\begin{array}{l}\alpha_1\beta_1+\alpha_1\beta_2+\alpha_1\beta_3+\cdots+\alpha_1\beta_n+\cdots\\ \alpha_2\beta_1+\alpha_2\beta_2+\alpha_2\beta_3+\cdots+\alpha_2\beta_n+\cdots\\ \alpha_3\beta_1+\alpha_3\beta_2+\alpha_3\beta_3+\cdots+\alpha_3\beta_n+\cdots\\ \cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\\ \alpha_n\beta_1+\alpha_n\beta_2+\alpha_n\beta_3+\cdots+\alpha_n\beta_n+\cdots\\ \cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\end{array}\right\} \qquad (6)$$

The theorem announces that if we make up a series by selecting the terms of this scheme in any manner so that each term appears in the resulting series, then that series converges absolutely and has the limit $\alpha\beta$.

Suppose we arrange the terms of the foregoing array into a series in the following order:

$$\alpha_1\beta_1+\alpha_1\beta_2+\alpha_2\beta_2+\alpha_2\beta_1+\alpha_1\beta_3+\cdots. \qquad (7)$$

The order of selection of the terms of the series follows the boundaries of successive squares from top to bottom and then from right to left as indicated in (6). We shall now show that series (7) converges absolutely. Put as before

$$|\alpha_n|=\rho_n, \qquad |\beta_n|=r_n,$$

and let the series $\Sigma\rho_n$, Σr_n converge respectively to the limits ρ and r. We may write for convenience

$$S_n'=\rho_1+\rho_2+\cdots+\rho_n,$$
$$T_n'=r_1+r_2+\cdots+r_n.$$

It is proposed to show that the series of moduli obtained from (7), namely

$$\rho_1r_1+\rho_1r_2+\rho_2r_2+\rho_2r_1+\rho_1r_3+\cdots, \qquad (8)$$

converges. Let M_k denote the sum of the first k terms of this series; then if we have

$$n^2\leqq k\leqq(n+1)^2,$$

it follows that

$$S_n'T_n'\leqq M_k\leqq S'_{n+1}T'_{n+1}.$$

But since both $\Sigma\alpha_n$, $\Sigma\beta_n$ converge absolutely it follows that

$$\underset{n=\infty}{L}S_n'T_n'=\underset{n=\infty}{L}S'_{n+1}T'_{n+1}=\underset{n=\infty}{L}S_n'\underset{n=\infty}{L}T_n'=\rho r.$$

Consequently, we have,* since k increases with n,

$$\underset{k=\infty}{L} M_k = \rho r$$

and series (7) converges absolutely.

Since the series (7) converges absolutely its terms may be grouped in any manner desired without affecting the limiting value of the series. Suppose we collect the terms of (7) into groups as follows:

$$\begin{aligned} &\alpha_1\beta_1 + (\alpha_1\beta_2 + \alpha_2\beta_2 + \alpha_2\beta_1) + \cdots \\ &+ (\alpha_1\beta_n + \cdots + \alpha_n\beta_n + \cdots + \alpha_n\beta_1) + \cdots . \end{aligned} \tag{9}$$

For convenience put

$$p_n = (\alpha_1\beta_n + \cdots + \alpha_n\beta_n + \cdots + \alpha_n\beta_1)$$

and

$$\begin{aligned} P_n &= p_1 + p_2 + \cdots + p_n, \\ S_n &= \alpha_1 + \alpha_2 + \cdots + \alpha_n, \\ T_n &= \beta_1 + \beta_2 + \cdots + \beta_n. \end{aligned}$$

We have then

$$P_n = S_n T_n.$$

Upon passing to the limit as n becomes infinite, we obtain

$$\underset{n=\infty}{L} P_n = \underset{n=\infty}{L} S_n T_n = \underset{n=\infty}{L} S_n \underset{n=\infty}{L} T_n = \alpha\beta.$$

Since the limit of the series (9) is equal to that of (7), it follows that the latter series not only converges absolutely but has the limiting value $\alpha\beta$.

The series (7) was obtained by selecting its terms from (6) in a particular manner. It has been shown, however, that (7) converges absolutely and consequently by Theorem VI, Art. 42, its terms may be rearranged and grouped in any manner desired. It follows then that whatever is the manner of arranging the terms of (6) into a series the limit is still $\alpha\beta$ and the resulting series converges absolutely. It is often convenient to select the terms of our series by taking the terms of (6) along diagonal lines as follows:

$$\alpha_1\beta_1 + (\alpha_1\beta_2 + \alpha_2\beta_1) + \cdots + (\alpha_1\beta_n + \alpha_2\beta_{n-1} + \cdots + \alpha_n\beta_1) + \cdots \tag{10}$$

In this arrangement of the product, each group contains all of the terms $\alpha_r\beta_s$ for which $r + s$ has the same value. From the foregoing discussion, it follows that this series likewise converges absolutely and has the limit $\alpha\beta$.

* See Townsend and Goodenough, *First Course in Calculus*, p. 24, Theorem VI.

Ex. 3. Given the two series,

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots,$$

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots,$$

to show that $2 \sin z \cos z = \sin 2z$.

We have

$$\begin{aligned}
2 \sin z \cos z = 2 \Big\{ & z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots \\
& - \frac{z^3}{2!} + \frac{z^5}{2!3!} - \frac{z^7}{2!5!} - \cdots \\
& + \frac{z^5}{4!} - \frac{z^7}{4!3!} + \cdots \\
& - \frac{z^7}{6!} + \cdots \\
& + \cdots\cdots\cdots\cdots\cdots \\
= {} & 2z - 2\left\{\frac{z^3}{3!} + \frac{z^3}{2!}\right\} + 2\left\{\frac{z^5}{5!} + \frac{z^5}{3!2!} + \frac{z^5}{4!}\right\} - \cdots \\
= {} & 2z - \frac{8z^3}{3!} + \frac{32z^5}{5!} - \cdots \\
= {} & (2z) - \frac{(2z)^3}{3!} + \frac{(2z)^5}{5!} - \cdots \\
= {} & \sin 2z.
\end{aligned}$$

Mertens has shown * that the series (10) converges to the limit $\alpha\beta$ when only one of the given series $\Sigma\alpha_n$, $\Sigma\beta_n$ converges absolutely. It may be shown † that when both of the series $\Sigma\alpha_n$, $\Sigma\beta_n$ converge conditionally then if the series (10) converges at all, it has the limiting value $\alpha\beta$. In neither of these cases, however, is the product necessarily an absolutely convergent series. On the other hand, the series (10) may still be absolutely convergent in particular cases when one of the series $\Sigma\alpha_n$, $\Sigma\beta_n$ is absolutely convergent and the other conditionally convergent, or even divergent; also when one of these series is conditionally convergent and the other divergent; and finally, when both are conditionally convergent or both divergent.‡

As division is the inverse operation of multiplication, the results of Theorem III may be used in determining the conditions under which we may divide one series by another. The resulting theorem may be stated as follows:

* See *Jour. für d. reine u. angew. Math.*, Vol. 79, p. 182; also Whittaker, *Modern Analysis*, Art. 19.

† See Whittaker, *Modern Analysis*, Art. 20.

‡ See Cajori, *Bulletin Amer. Math. Soc.*, Jan. 1903.

THEOREM IV. *Given the series $\Sigma\alpha_n$ and $\Sigma\beta_n$ having the limits α and β, respectively. Suppose $\Sigma\beta_n$ converges absolutely and that $\beta_1 \neq 0$. Then*

$$\frac{\alpha}{\beta} = \frac{\Sigma\alpha_n}{\Sigma\beta_n} = \lambda_1 + \lambda_2 + \cdots + \lambda_n + \cdots, \tag{11}$$

where

$$\lambda_n = \frac{\alpha_n - \lambda_1\beta_n - \lambda_2\beta_{n-1} - \cdots - \lambda_{n-1}\beta_2}{\beta_1},$$

provided the series $\Sigma\lambda_n$ converges absolutely.

In order that we may have the relation

$$\frac{\Sigma\alpha_n}{\Sigma\beta_n} = \Sigma\lambda_n,$$

the series $\Sigma\alpha_n$ must be the product of $\Sigma\beta_n$ and $\Sigma\lambda_n$. Since both of the series $\Sigma\beta_n$, $\Sigma\lambda_n$ converge absolutely, they may be multiplied term by term and the product written in the form

$$\lambda_1\beta_1 + (\lambda_1\beta_2 + \lambda_2\beta_1) + \cdots + (\lambda_1\beta_n + \lambda_2\beta_{n-1} + \cdots + \lambda_{n-1}\beta_2 + \lambda_n\beta_1) + \cdots.$$

If this series is to be identical with the series

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n + \cdots,$$

then the value of $\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots$ must be so determined that the corresponding terms of the two series are equal; that is, we must have

$$\begin{aligned}
\alpha_1 &= \lambda_1\beta_1, \\
\alpha_2 &= \lambda_1\beta_2 + \lambda_2\beta_1, \\
&\cdots\cdots \\
\alpha_n &= \lambda_1\beta_n + \lambda_2\beta_{n-1} + \cdots + \lambda_{n-1}\beta_2 + \lambda_n\beta_1, \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

whence we have, since $\beta_1 \neq 0$,

$$\begin{aligned}
\lambda_1 &= \frac{\alpha_1}{\beta_1}, \\
\lambda_2 &= \frac{\alpha_2 - \lambda_1\beta_2}{\beta_1}, \\
&\cdots\cdots \\
\lambda_n &= \frac{\alpha_n - \lambda_1\beta_n - \lambda_2\beta_{n-1} - \cdots - \lambda_{n-1}\beta_2}{\beta_1}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

This completes the demonstration of the theorem.

Ex. 4. Given the two series

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots,$$

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots;$$

find

$$\tan z = z + \frac{z^3}{3} + \frac{2z^5}{15} + \cdots.$$

We have, since $\beta_1 = 1$,

$$\lambda_1 = \alpha_1 = z,$$

$$\lambda_2 = \alpha_2 - \lambda_1\beta_2 = -\frac{z^3}{3!} - z\left(-\frac{z^2}{2!}\right) = \frac{z^3}{3},$$

$$\lambda_3 = \alpha_3 - \lambda_1\beta_3 - \lambda_2\beta_2 = \frac{z^5}{5!} - z\left(\frac{z^4}{4!}\right) - \frac{z^3}{3}\left(-\frac{z^2}{2!}\right) = \frac{2z^5}{15},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

44. Double series. Let us consider a doubly infinite array of complex elements of the form

$$\begin{array}{llllll} \alpha_{11} & \alpha_{12} & \alpha_{13} & \ldots & \alpha_{1n} & \ldots, \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \ldots & \alpha_{2n} & \ldots, \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \ldots & \alpha_{3n} & \ldots, \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha_{m1} & \alpha_{m2} & \alpha_{m3} & \ldots & \alpha_{mn} & \ldots, \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array} \tag{1}$$

where α_{mn} indicates the element in the m^{th} row and the n^{th} column. Each row of the array extends indefinitely to the right and each column extends indefinitely downwards. If the elements of this array are connected by plus signs, the result is called a **double series**, denoted by

$$\Sigma\alpha_{mn}. \tag{2}$$

Let S_{mn} denote the sum of the elements in the first m rows and n columns of the given array. If S_{mn} approaches some definite limit α as m and n become infinite independently of each other and if this limit is independent of the manner in which m and n become infinite, that is if

$$\underset{\substack{m=\infty \\ n=\infty}}{L} S_{mn} = \alpha, \tag{3}$$

then the double series (2) is said to **converge** to the limit α, and α is called the **sum** of the series. If the limit (3) does not exist, the series is said to be divergent. If the series formed by taking the

moduli of the elements of (1) converges, then the given series is said to converge absolutely.

The product of two series $\Sigma\alpha_n \cdot \Sigma\beta_n$, already discussed, may be exhibited as a double series. We need merely connect the various rows of (6) in the previous article with the plus sign. The double series was in that case converted into a simple series by adding the elements of the array along the two sides of successive squares. The order of the terms in the resulting simple series is that of (7), Art. 43. If this series converges absolutely, then by Theorem V, Art. 42, it converges independently of the order of the terms. The terms can therefore be so rearranged as to give various simple series equivalent to the given double series, every such series converging to the same limit, namely the limit of the double series. We may say, moreover, that if the moduli of the elements of the double series $\Sigma\alpha_{mn}$ can be arranged in any one way so as to form a simple convergent series, then the double series $\Sigma\alpha_{mn}$ **converges absolutely.** Such a double series may then be converted into a simple series in any arbitrary manner by means of which there is set up a one-to-one correspondence between the totality of elements of the double series and the positive integers representing their order as terms in the simple series.

The following generalization of Theorem VI, Art. 42, makes the use of absolutely convergent double series of advantage in some discussions.

THEOREM. *If a double series of complex terms converges absolutely, it may be evaluated by rows or columns.*

Let $\Sigma\alpha_{mn}$ be the given double series whose terms are the elements in the array (1). Suppose this series converges absolutely and has the limit α. We are to show that the limits of the series constituting the various rows exist and that the sum of these limits forms a series having α as a limit. Furthermore, we are to show that the limits of the series constituting the various columns exist and that the sum of these limits likewise forms a series having the limit α of the double series.

Let the series of moduli of the terms of $\Sigma\alpha_{mn}$, namely $\Sigma\rho_{mn}$, converge to the limit ρ. The series formed by taking the moduli of the elements of any row must converge; for, the sum of a finite number of such terms increases with n and is always less than ρ. We shall denote the limiting values of the successive rows of moduli by ρ_1,

$\rho_2, \ldots, \rho_m, \ldots$. Since each row of the series of moduli converges it follows from Theorem III, Art. 42, that each row of the series $\Sigma\alpha_{mn}$ converges. Let the successive rows converge respectively to the limits $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_m, \ldots$, and let S_m denote the sum of the first m terms of this series.

We shall now show that the series $\Sigma\alpha_m$ converges to the limit α. This series will converge if the series of moduli $\Sigma \mid \alpha_m \mid$ converges. We have for all values of m

$$| \alpha_m | \leqq \rho_m,$$

and since the sum

$$\rho_1 + \rho_2 + \cdots + \rho_m$$

increases with m and is always less than ρ, the series $\Sigma\rho_m$ converges, and by comparison $\Sigma \mid \alpha_m \mid$ converges. Therefore $\Sigma\alpha_m$ converges, say to α'.

It remains to show that $\alpha' = \alpha$. Denote by $R_1, R_2, \ldots, R_m$, respectively, the absolute values of the remainders after the first n terms of the first m rows of the given double series $\Sigma\alpha_{mn}$; that is, let

$$R_i = | \alpha_{i,\, n+1} + \alpha_{i,\, n+2} + \cdots |, \quad i = 1, 2, \cdots, \quad n > m.$$

For an arbitrarily chosen ϵ, we may now select a number N_m depending upon m, such that for $n > N_m$ each of the numbers R_1, $R_2, \ldots, R_m$ is less than $\frac{\epsilon}{m}\cdot$ We have then

$$| S_{mn} - S_m | \leqq R_1 + R_2 + \cdots + R_m < \epsilon, \quad n > N_m. \tag{4}$$

However, since the series $\Sigma\alpha_m$ converges to the limit α', we have for some value m_1,

$$| S_m - \alpha' | < \epsilon, \qquad m > m_1. \tag{5}$$

Combining (4) and (5), we obtain

$$| S_{mn} - \alpha' | < 2\epsilon, \qquad m > m_1, \qquad n > N_m. \tag{6}$$

Since the limit $\underset{\substack{m=\infty \\ n=\infty}}{L} S_{mn}$ exists, we have from (6)

$$\underset{\substack{m=\infty \\ n=\infty}}{L} S_{mn} = \alpha'.$$

By hypothesis S_{mn} has the limit α as m and n become infinite. Hence, α' must be identical with α and the series $\Sigma\alpha_m$ converges to α as the theorem requires.

A similar argument shows that when the given double series Σa_{mn} is evaluated by columns, the limiting value is again a, the sum of the given series.

Ex. Consider the following double series, which is of importance in the Weierstrassian theory of elliptic functions:

$$\begin{array}{cccccccccccc}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdots + & \dfrac{1}{(-4\,\omega_1+4\,\omega_3)^3} & + & \dfrac{1}{(-2\,\omega_1+4\,\omega_3)^3} & + & \dfrac{1}{(4\,\omega_3)^3} & + & \dfrac{1}{(2\,\omega_1+4\,\omega_3)^3} & + & \dfrac{1}{(4\,\omega_1+4\,\omega_3)^3} & + \\
\cdots + & \dfrac{1}{(-4\,\omega_1+2\,\omega_3)^3} & + & \dfrac{1}{(-2\,\omega_1+2\,\omega_3)^3} & + & \dfrac{1}{(2\,\omega_3)^3} & + & \dfrac{1}{(2\,\omega_1+2\,\omega_3)^3} & + & \dfrac{1}{(4\,\omega_1+2\,\omega_3)^3} & + \\
\cdots + & \dfrac{1}{(-4\,\omega_1)^3} & + & \dfrac{1}{(-2\,\omega_1)^3} & + & 0 & + & \dfrac{1}{(2\,\omega_1)^3} & + & \dfrac{1}{(4\,\omega_1)^3} & + \\
\cdots + & \dfrac{1}{(-4\,\omega_1-2\,\omega_3)^3} & + & \dfrac{1}{(-2\,\omega_1-2\,\omega_3)^3} & + & \dfrac{1}{(-2\,\omega_3)^3} & + & \dfrac{1}{(2\,\omega_1-2\,\omega_3)^3} & + & \dfrac{1}{(4\,\omega_1-2\,\omega_3)^3} & + \\
\cdots + & \dfrac{1}{(-4\,\omega_1-4\,\omega_3)^3} & + & \dfrac{1}{(-2\,\omega_1-4\,\omega_3)^3} & + & \dfrac{1}{(-4\,\omega_3)^3} & + & \dfrac{1}{(2\,\omega_1-4\,\omega_3)^3} & + & \dfrac{1}{(4\,\omega_1-4\,\omega_3)^3} & + \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{array}$$

where ω_1 and ω_3 are any two complex numbers subject only to the restriction that the real part of $\left(\dfrac{1}{i}\,\dfrac{\omega_3}{\omega_1}\right)$ shall be greater than zero. The points

$$\Omega = 2\,m_1\omega_1 + 2\,m_3\omega_3,$$

where m_1, m_3 are integers, lie on a network of parallel lines covering the entire complex plane.

The series under consideration is

$$\Sigma'\,\frac{1}{\Omega^3},$$

where Σ' denotes the sum for all values of Ω except the value for which m_1, m_3 are both equal to zero.

This double series may be converted into a simple series and its absolute convergence established, by selecting the points Ω in order along the sides of the successive parallelograms as indicated in Fig. 81.

Let these successive parallelograms be denoted by

$$P_1,\ P_2,\ P_3,\ \ldots,\ P_n,\ \ldots.$$

Let l be the least and L the greatest distance of any point of P_1 from the origin. On the perimeter of P_1 are 8 points Ω, such that for each point

$$\left|\frac{1}{\Omega^3}\right| \leqq \frac{1}{l^3}.$$

On the perimeter of P_n are $8\,n$ points Ω, such that for each point

$$\left|\frac{1}{\Omega^3}\right| \leqq \frac{1}{(nl)^3}.$$

Hence, we have

$$\sum{}' \left| \frac{1}{\Omega^3} \right| \leqq \sum_{n=1}^{\infty} \frac{8n}{(nl)^3} = \frac{8}{l^3} \left\{ 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots \right\}.$$

Since the terms of the given series are less in absolute value than the corresponding terms of the well-known convergent series included in the braces multiplied by a constant $\frac{8}{l^3}$, it follows that the given double series converges absolutely.

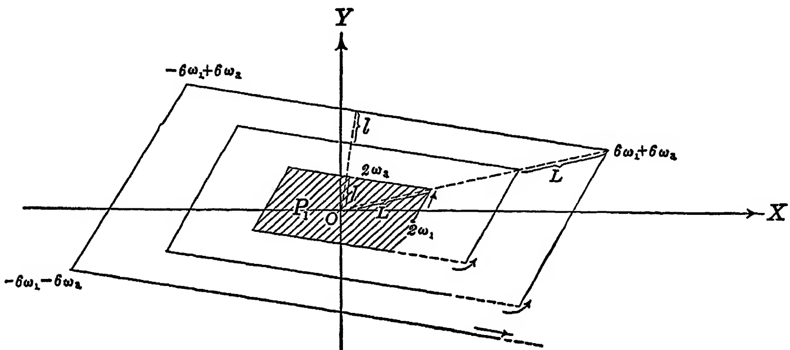

FIG. 81.

It may be remarked that, on the other hand, the double series

$$\sum{}' \left| \frac{1}{\Omega^2} \right|$$

does not converge; for, on the perimeter of P_n are $8n$ points Ω, such that for each point we have

$$\left| \frac{1}{\Omega^2} \right| \geqq \frac{1}{(nL)^2}.$$

Hence, we obtain

$$\sum{}' \left| \frac{1}{\Omega^2} \right| \geqq \sum_{n=1}^{\infty} \frac{8n}{(nL)^2} = \frac{8}{L^2} \left\{ 1 + \frac{1}{2} + \frac{1}{3} + \cdots \right\},$$

where the series inclosed in the braces is known to be divergent.

45. Uniform convergence. Suppose we have given a series of functions

$$u_1(z) + u_2(z) + \cdots + u_n(z) + \cdots,$$

and suppose this series converges for all values of z in a given region S which may or may not be closed. We shall speak of S as the

region of convergence. In this region the series then defines a function, and we may write

$$f(z) = \underset{n=\infty}{L}\, S_n(z),$$

where $S_n(z) \equiv \sum_{1}^{n} u_n(z)$. We may also write

$$f(z) = S_n(z) + R_n(z),$$

where $R_n(z)$ represents the remainder after the first n terms.

It is often true that the given series will converge more rapidly in the neighborhood of certain points than in the neighborhood of others. Let z_1 be some point in S. Since the series converges for this value of z, it is possible to find, for an arbitrarily small positive number ϵ, a positive integer, say m_1, such that for all values of $n > m_1$, we have

$$| f(z_1) - S_n(z_1) | \equiv | R_n(z_1) | < \epsilon.$$

If the value of ϵ is kept constant, it will in general be necessary to select a new integer m_2 if z_1 be replaced by some other value z_2 of S. If, in the selection of m, the least integer that will answer the purpose is taken, then with each point z there is associated a particular integer m, namely the first integer for which we have $| R_n(z) | < \epsilon$, where $n > m$. We may then write $m(z)$ as a function of z. Consider now the totality of all the values of m corresponding to the points of the region S. These values of m may or may not have a finite upper limit. In case a finite upper limit exists, we may say that the series converges uniformly in the region S. Denoting the upper limit of $m(z)$ by M, then for a given ϵ any integer $m > M$ may be associated equally well with each value of z. The definition of **uniform convergence** may now be stated as follows:

The given series is said to converge uniformly in a given region S, closed or not, if corresponding to an arbitrarily small positive number ϵ it is possible to find an integer m, which is independent of z, such that for all values of $n > m$ we have simultaneously for all values of z, in the region S,

$$| f(z) - S_n(z) | \equiv | R_n(z) | < \epsilon.$$

The following example furnishes an illustration of regions of uniform and of non-uniform convergence of a series of functions.

Ex. 4. Given the series

$$z^2 + \frac{z^2}{1+z^2} + \frac{z^2}{(1+z^2)^2} + \cdots + \frac{z^2}{(1+z^2)^{n-1}} + \cdots.$$

This is a geometric series; hence, we have

$$S_n(z) = 1 + z^2 - \frac{1}{(1+z^2)^{n-1}}.$$

We shall consider the character of the convergence of this series for finite values of $z = \rho(\cos\theta + i\sin\theta)$ in the region defined by the inequalities

$$\rho \geqq 0, \qquad -\frac{\pi}{4} \leqq \theta \leqq \frac{\pi}{4}.$$

In this region the series converges and defines the function

$$\begin{aligned} f(z) &= 1 + z^2, &&\text{for } z \neq 0,\\ &= 0, &&\text{for } z = 0. \end{aligned}$$

While the given series converges in the region indicated, that region is not to be understood as the whole of the region of convergence. The remainder after the first n terms is, in the region under consideration,

$$\begin{aligned} R_n(z) &= \frac{1}{(1+z^2)^{n-1}}, &&\text{for } z \neq 0,\\ &= 0, &&\text{for } z = 0. \end{aligned}$$

For an arbitrarily small ϵ, there corresponds to each value of z an integer m such that

$$|R_n(z)| < \epsilon, \qquad n > m.$$

But there is no integer m, however large it may be chosen, that answers this purpose simultaneously for all values of z; for, suppose we take $m = G$, chosen as large as we please, then for $n_1 > G$ we may always find values of z such that

$$|R_{n_1}(z)| \equiv \left|\frac{1}{(1+z^2)^{n_1-1}}\right| > \epsilon.$$

We need only choose z so that $|z|$ is sufficiently small. It follows then that the given series is non-uniformly convergent in the region selected.

Suppose we now restrict the region by excluding a region about the origin; that is, let

$$\rho \geqq r, \quad \text{where } 0 < r < 1.$$

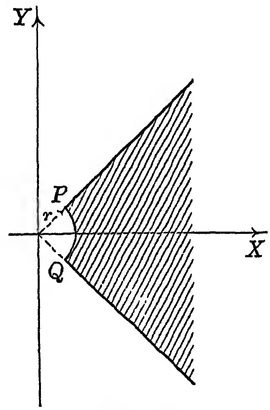

FIG. 82.

The lower limit of $|1 + z^2|$ in the new region is then $\sqrt{1 + r^4}$, which is the value of $|1 + z^2|$ when z is at P or Q, Fig. 82. Hence, the upper limit of

$$|R_n(z)| \equiv \left|\frac{1}{(1+z^2)^{n-1}}\right|$$

for a given n is $\dfrac{1}{(1+r^4)^{\frac{n-1}{2}}}$. In order to determine a value of m such that for every z of the region we have

$$|R_n(z)| < \epsilon, \qquad n > m,$$

put

$$\frac{1}{(1+r^4)^{\frac{m'-1}{2}}} = \epsilon,$$

whence

$$m' = 1 - \frac{2\log\epsilon}{\log(1+r^4)}.$$

Then for all values of $n > m > m'$, we have

$$|R_n z| \equiv \left|\frac{1}{(1+z^2)^{n-1}}\right| < \epsilon,$$

irrespective of the value of z in the finite region where

$$\rho \geqq r, \qquad \frac{-\pi}{4} \leqq \theta \leqq \frac{\pi}{4}.$$

In this region then the given series converges uniformly.

It will be observed that the region of uniform convergence may not coincide with the region of convergence. In fact, it is frequently more restricted than the region of convergence. As a convenient test for uniform convergence, we have the following theorem, due to Weierstrass.

THEOREM I. *Given the series*

$$u_1(z) + u_2(z) + \cdots + u_n(z) + \cdots.$$

If for all values of z in a given region S, closed or not, we have for all values of n

$$|u_n(z)| \leqq M_n,$$

where ΣM_n is a convergent series of positive constants, then $\Sigma u_n(z)$ converges absolutely and uniformly in S.

The absolute convergence of Σu_n follows at once from the foregoing discussion of absolute convergence, since by the conditions of the theorem we have $|u_n(z)| \leqq M_n$ and ΣM_n is convergent.

The uniform convergence of the series may be established as follows. Since the series ΣM_n is convergent, we can find a number m such that

$$\sum_{n=m+1}^{\infty} M_n < \epsilon.$$

Moreover, we may write

$$\left|\sum_{m+1}^{m+p} u_n(z)\right| \leqq \sum_{m+1}^{m+p} M_n < \sum_{m+1}^{\infty} M_n, \quad p = 1, 2, 3, \ldots .$$

Consequently, we have

$$|R_m(z)| \equiv \underset{p=\infty}{L} \left|\sum_{m+1}^{m+p} u_n(z)\right| < \epsilon.$$

Since this relation exists independently of the value z may take in the region S, the condition set forth in the definition of uniform convergence is satisfied.

It is to be noted that the foregoing theorem gives a sufficient but not a necessary condition for uniform convergence. Other and more delicate tests for uniform convergence may be made to apply to series of functions of a complex variable,* but the one given is sufficient to test those series to be considered in the present volume.

Uniform convergence gives a useful criterion for the continuity of the function defined by a convergent series of functions. This criterion may be stated as follows:

THEOREM II. *Given the function $f(z)$ defined by the convergent series*

$$u_1(z) + u_2(z) + \cdots + u_n(z) + \cdots .$$

If $u_n(z)$ is continuous and the series converges uniformly in a region S closed or not, then $f(z)$ is continuous in S.

We may write the given function in the form

$$f(z) = S_n(z) + R_n(z).$$

Since the series converges uniformly in a given region S, it follows that for any value z_0 in S we have, for $n > m$,

$$|R_n(z_0)| < \epsilon. \tag{1}$$

Suppose z takes an increment $\Delta_1 z$ such that $z_0 + \Delta_1 z$ lies in S. We then have

$$|R_n(z_0 + \Delta_1 z)| < \epsilon,$$

whence

$$|\Delta R_n(z_0)| = |R_n(z_0 + \Delta_1 z) - R_n(z_0)| < 2\epsilon. \tag{2}$$

Since $S_n(z)$ is continuous in z for all finite values of n, we have

$$|\Delta S_n(z_0)| = |S_n(z_0 + \Delta_2 z) - S_n(z_0)| < \epsilon. \tag{3}$$

* See Bromwich, *Theory of Infinite Series*, Art. 81.

By combining (2) and (3), we obtain

$$\begin{aligned} |\,\Delta f(z_0)\,| &= |\,\Delta R_n(z_0) + \Delta S_n(z_0)\,| \\ &\leqq |\,\Delta R_n(z_0)\,| + |\,\Delta S_n(z_0)\,| \\ &< 3\,\epsilon, \qquad |\,\Delta z\,| \leqq |\,\Delta_2 z\,| \leqq |\,\Delta_1 z\,|, \end{aligned}$$

for all values of Δz equal to the smaller of the increments $\Delta_1 z$, $\Delta_2 z$. Hence, since $3\,\epsilon$ is arbitrarily small, $|\,\Delta f(z_0)\,|$ is arbitrarily small and $f(z)$ is continuous at any point z_0 in the region S of uniform convergence.

46. Integration and differentiation of series. We shall frequently have occasion to integrate or differentiate a series term by term. The question arises whether the resulting series represents the integral or derivative, as the case may be, of the function defined by the given series. Suppose a function $f(z)$ is defined by the relation

$$f(z) = u_1(z) + u_2(z) + \cdots + u_n(z) + \cdots,$$

where the u's are continuous functions in a region S within which the given curve C lies. Denote by $S_n(z)$ the sum of the first n terms of this series. The integral of the function along the path C, if it exists, may then be written

$$\int_C f(z)\,dz = \int_C \underset{n=\infty}{L}\, S_n(z)\,dz. \tag{1}$$

For any definite value of n, we may write

$$\sum_1^n \int_C u_n(z)\,dz = \int_C \sum_1^n u_n(z)\,dz = \int_C S_n(z)\,dz.$$

Consequently, the result of term by term integration of the series defining $f(z)$ may be written

$$\underset{n=\infty}{L} \int_C S_n(z)\,dz. \tag{2}$$

It can not be assumed that the two results (1), (2) are equal. The following example furnishes an illustration where (1) and (2) are not equal.

Ex. 1. Given the series, the sum of whose first n terms is $S_n(z) = nze^{-nz^2}$. Consider the term by term integration of this series where the path C of integration is the X-axis from 0 to any point β, $0 < \beta < 1$.

The series converges for real values of z. The integral along the axis of reals is an ordinary definite integral for real values of z. We have then

$$\int_0^\beta f(z)\,dz = \int_0^\beta \underset{n=\infty}{L}\, nze^{-nz^2}\,dz = \int_0^\beta 0\,dz = 0.$$

On the other hand, we have

$$\underset{n=\infty}{L} \int_0^\beta S_n(z)\, dz = \underset{n=\infty}{L} \int_0^\beta nze^{-nz^2}\, dz = \underset{n=\infty}{L}\ \frac{1}{2}(1 - e^{-n\beta^2}) = \frac{1}{2}.$$

Hence in this case we can not integrate the given series term by term.

We shall now set up a condition by means of uniform convergence that will be sufficient for the equality of (1) and (2). This condition may be stated as follows:

THEOREM I. *Let $f(z)$ be defined by the convergent series*

$$u_1(z) + u_2(z) + \cdots + u_n(z) + \cdots,$$

where $u_n(z)$ is a continuous function for values of z along an ordinary curve C. If the series converges uniformly along C, we may integrate the series term by term, thus obtaining

$$\int_C f(z)\, dz = \int_C u_1(z)\, dz + \int_C u_2(z)\, dz + \cdots + \int_C u_n(z)\, dz + \cdots. \quad (3)$$

Each term of the series is continuous and hence the integral $\int_C u_n(z)\, dz$ exists. Moreover, since the series converges uniformly, the function $f(z)$ is a continuous function and the integral $\int_C f(z)\, dz$ also exists. We shall now show that the relation given in (3) holds.

We may write

$$f(z) = u_1(z) + u_2(z) + \cdots + u_n(z) + R_n(z), \quad (4)$$

where $R_n(z)$ denotes the remainder of the given series after the first n terms. By formula 3, Art. 17, we have

$$\begin{aligned}\int_C f(z)\, dz &= \int_C \{u_1(z) + u_2(z) + \cdots + u_n(z) + R_n(z)\}dz \\ &= \int_C u_1(z)\, dz + \int_C u_2(z)\, dz + \cdots + \int_C u_n(z)\, dz + \int_C R_n(z)\, dz. \quad (5)\end{aligned}$$

For n sufficiently large, say $n > m$, the integral $\int_C R_n(z)\, dz$ becomes arbitrarily small. For, we have

$$\left|\int_C R_n(z)\, dz\right| \leqq \int_C |R_n(z)| \cdot |dz|. \quad (6)$$

As the series converges uniformly, we have for all values of z along C

$$|R_n(z)| < \epsilon, \qquad n > m.$$

Hence, from (6) we obtain

$$\left| \int_C R_n(z)\, dz \right| < \epsilon \int_C |\, dz \,|, \qquad n > m$$
$$= \epsilon \cdot L,$$

where L denotes the length of the path C of integration and is therefore finite. Since $\epsilon \cdot L$ is arbitrarily small, we have

$$\underset{n=\infty}{L} \int_C R_n(z)\, dz = 0.$$

Consequently, when n is allowed to increase without limit, we obtain from (5) the relation given in the theorem.

It is necessary also to set up some criterion for the differentiation of a series term by term; for, it can not be assumed that the series formed by differentiating the various terms of a given convergent series is equal to the derivative of the function defined by that series. The following example furnishes an illustration.

Ex. 2. Given the series

$$\sin z - \frac{\sin 2z}{2} + \frac{\sin 3z}{3} - \frac{\sin 4z}{4} + \cdots.$$

The series converges for real values of the variable, and defines the function $\frac{z}{2}$ for values of z lying between $-\pi$ and π. Differentiating the series term by term, we get

$$\cos z - \cos 2z + \cos 3z - \cos 4z + \cdots.$$

This series of derivatives does not represent the derivative of $\frac{z}{2}$; for, it does not even converge. For, in order that a series shall converge, we must have, as we have seen, the limit of the n^{th} term equal to zero. However, in the case under consideration, the limit $\underset{n=\infty}{L} |\cos nz|$ does not even exist for $z \neq 0$.

If the terms of the given series are holomorphic in a given region S, we have the following theorem, which furnishes a convenient criterion for differentiating or integrating term by term such series as we shall have occasion to consider. It also furnishes a condition that the function defined by the series shall be holomorphic in S.

THEOREM II. *Let $f(z)$ be defined by the convergent series*

$$u_1(z) + u_2(z) + \cdots + u_n(z) + \cdots,$$

where $u_n(z)$ is holomorphic in a region S. If this series converges uniformly in every simply connected closed region lying wholly in S and bounded by an ordinary curve C, then the series may be integrated or differentiated term by term for values of z in S. Moreover, $f(z)$ is holomorphic in S.

Since the given series converges uniformly and each term is continuous, it follows from Theorem I that it may be integrated term by term along any ordinary curve C lying in S. It is to be noted that as $u_n(z)$ is also holomorphic in S, the integral of each term of the series is zero, since C is the complete boundary of a simply connected closed region lying wholly in S. We have then

$$\int_C f(z)\,dz = 0.$$

Consequently, by Theorem IV, Art. 20, $f(z)$ is holomorphic in the given region S.

To show that the given series may be differentiated term by term, we proceed as follows. Consider the series

$$f(t) = u_1(t) + u_2(t) + \cdots + u_n(t) + \cdots, \tag{7}$$

where t takes values along the closed curve C. This series converges uniformly, a property that is not destroyed by multiplying the terms of the series by the factor $\frac{1}{2\pi i(t-z)^2}$, where z is any point within C. We have then

$$\frac{1}{2\pi i}\frac{f(t)}{(t-z)^2} = \frac{1}{2\pi i}\frac{u_1(t)}{(t-z)^2} + \frac{1}{2\pi i}\frac{u_2(t)}{(t-z)^2} + \cdots + \frac{1}{2\pi i}\frac{u_n(t)}{(t-z)^2} + \cdots.$$

Integrating term by term, we obtain

$$\frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{(t-z)^2} = \frac{1}{2\pi i}\int_C \frac{u_1(t)\,dt}{(t-z)^2} + \frac{1}{2\pi i}\int_C \frac{u_2(t)\,dt}{(t-z)^2} + \cdots + \frac{1}{2\pi i}\int_C \frac{u_n(t)\,dt}{(t-z)^2} + \cdots.$$

From Art. 20, it will be seen that the terms of this series of integrals are the first derivatives of the terms of the given series. We have therefore

$$f'(z) = u_1'(z) + u_2'(z) + \cdots + u_2'(z) + \cdots \tag{8}$$

for any value of z within C. But C is any closed curve in S; hence (8) holds for any values of z in S as stated in the theorem.

Ex. 3. Given the series

$$\frac{1}{1\cdot 3\cdot 5}+\frac{z}{3\cdot 5\cdot 7}+\frac{z^2}{5\cdot 7\cdot 9}+\cdots+\frac{z^{n-1}}{(2n-1)(2n+1)(2n+3)}+\cdots.$$

This series converges uniformly in any region bounded by a circle about the origin, which is situated within the unit circle. It represents the function * which is given, except for $z=0$, by the expression

$$\frac{1}{8z^2}\left\{\frac{(1-z)^2}{2\sqrt{z}}\log\frac{1+\sqrt{z}}{1-\sqrt{z}}+\frac{5}{3}z-1\right\}.$$

The indefinite integral of this function is readily found by integrating the given series term by term, thus obtaining

$$C+\frac{z}{1\cdot 3\cdot 5}+\frac{z^2}{2\cdot 3\cdot 5\cdot 7}+\frac{z^3}{3\cdot 5\cdot 7\cdot 9}+\cdots+\frac{z^n}{n(2n-1)(2n+1)(2n+3)}+\cdots.$$

Ex. 4. Given the series

$$z-\frac{z^3}{3!}+\frac{z^5}{5!}-\frac{z^7}{7!}+\cdots.$$

. . . .

This series converges uniformly in any finite region. The derived series is

$$1-\frac{z^2}{2!}+\frac{z^4}{4!}-\frac{z^6}{6!}+\cdots.$$

Consequently, the second series represents the derivative of the function defined by the first.

47. Power series. A series of the form

$$\alpha_0+\alpha_1 z+\alpha_2 z^2+\cdots+\alpha_n z^n+\cdots,$$

where n is a positive integer and

$$\alpha_n=a_n+ib_n=\rho_n(\cos\theta_n+i\sin\theta_n),$$
$$z=x+iy=r(\cos\phi+i\sin\phi),$$

is called a **power series** of complex terms. A more general form of a power series may be written

$$\alpha_0+\alpha_1(z-z_0)+\alpha_2(z-z_0)^2+\cdots+\alpha_n(z-z_0)^n+\cdots.$$

To distinguish the two types, we may speak of the second as a power series in $(z-z_0)$. For the sake of simplicity we shall confine our discussions for the most part to power series of the first type. In doing so there is no loss of generality, as power series in $z-z_0$ may be readily transformed into series of this type. Because of their importance, we shall consider some of the special properties of the power series. Among these properties is the following:

* Schlömilch, *Übungsbuch zum Studium der Höheren Analysis*, 4th Ed., Vol. 2, p. 239.

THEOREM I. *If for some positive number G we have for all values of n*

$$|\alpha_n| \cdot |z_0^n| \leqq G,$$

where z_0 is a constant value of z, then $\Sigma\alpha_n z^n$ converges absolutely for all values of z for which $|z| < |z_0|$.

Denoting the modulus of z_0 by r_0, we have by the conditions of the theorem

$$\rho_n r_0^n \leqq G.$$

The series of absolute values may be written

$$\rho_0 + \rho_1 r + \cdots + \rho_n r^n + \cdots = \rho_0 + \rho_1\left(\frac{r}{r_0}\right)r_0 + \cdots + \rho_n\left(\frac{r}{r_0}\right)^n r_0^n + \cdots$$

$$\leqq G\left\{1 + \frac{r}{r_0} + \cdots + \left(\frac{r}{r_0}\right)^n + \cdots\right\}.$$

The series within the braces converges to the limiting value $\dfrac{1}{1 - \dfrac{r}{r_0}}$ if we have $r < r_0$. Consequently, for $|z| < |z_0|$ the series of moduli converges, and hence the given series $\Sigma\alpha_n z^n$ converges absolutely as the theorem states.

Ex. 1. Test the convergence of the series

$$\sin z + \frac{1}{2}\sin^2 z + \frac{1}{3}\sin^3 z + \cdots + \frac{1}{n}\sin^n z + \cdots.$$

The given series is a power series in $\sin z$. If we put

$$w = \sin z,$$

we have

$$w + \frac{w^2}{2} + \frac{w^3}{3} + \cdots + \frac{w^n}{n} + \cdots.$$

This series converges for $|w| < 1$; for, we have then

$$\left|\frac{w^n}{n}\right| < 1$$

for all values of n. By the foregoing theorem the series converges absolutely within the circle of unit radius about the origin in the W-plane.

To find the region in the Z-plane within which the given series converges, it is necessary to map upon the Z-plane the circle about the origin in the W-plane having the radius 1 by means of the relation

$$w = \sin z.$$

We have then

$$u + iv = \sin(x + iy) = \sin x \cosh y + i \cos x \sinh y,$$

whence

$$u = \sin x \cosh y, \quad v = \cos x \sinh y.$$

The equation of the circle in the W-plane is

$$u^2 + v^2 = 1.$$

Substituting the values of u, v in terms of x, y, we get

$$\sin^2 x \cosh^2 y + \cos^2 x \sinh^2 y = 1,$$

which reduces to the form

$$\cosh 2y = \cos 2x + 2,$$

or

$$\sinh^2 y = \cos^2 x.$$

The portion of the fundamental region $-\frac{\pi}{2} < x \leqq \frac{\pi}{2}$ for $\sin z$ bounded by the curve given by this equation is shown in Fig. 83 and 84.

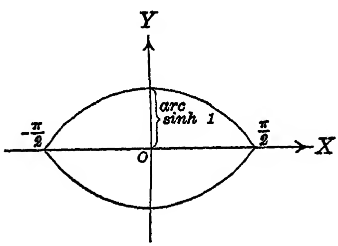

FIG. 83.

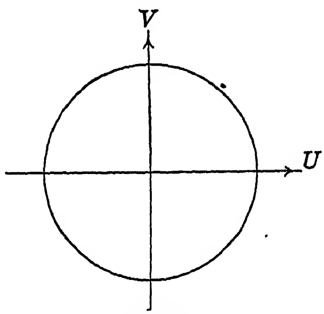

FIG. 84.

THEOREM II. *If the power series $\Sigma a_n z^n$ converges for $z = z_0$, it converges absolutely for all values of z for which $|z| < |z_0|$.*

This theorem follows as an immediate consequence of Theorem I. For if the given series converges for $z = z_0$, then there must exist some positive number G such that for all values of n

$$|a_n| \cdot |z_0^n| < G,$$

and consequently the series $\Sigma a_n z^n$ converges absolutely for values of z for which $|z| < |z_0|$ as the theorem requires.

We have also the following theorem.

THEOREM III. *If the power series $\Sigma a_n z^n$ is divergent for $z = z_1$, then it is divergent for all values of z for which $|z| > |z_1|$.*

By hypothesis the given series $\Sigma a_n z^n$ is divergent for $z = z_1$. It must then be divergent for all values of z where $|z| > |z_1|$; for, if

it is convergent for any such value of z, say z_2, where $|z_2| > |z_1|$, then it must, by Theorem II, be convergent for all values of z whose modulus is less than $|z_2|$ and therefore for $z = z_1$, which is a contradiction of the given hypothesis. From the contradiction the theorem follows.

Theorem II states that if a given series converges for $z = z_0$, then it converges within a circle about the origin having a radius equal to $|z_0|$; and Theorem III states that if it is divergent for $z = z_1$, then it is divergent for all values of z exterior to the circle about the origin whose radius is $|z_1|$. Nothing is said about the convergence of the series within the region between these two circles, or indeed upon the circles themselves, except at the points z_0 and z_1. The question presents itself as to whether it is possible to find a circle about the origin of radius R such that the given power series shall be convergent for all values of z where $|z| < R$ and divergent for all values of z where $|z| > R$.

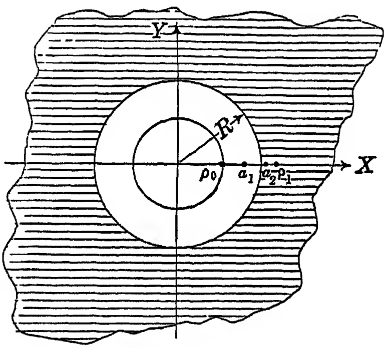

FIG. 85.

It may be shown as follows that such a circle of radius R always exists, where R may be zero, finite and different from zero, or infinite. Let as before z_0 be a point of convergence and z_1 a point of divergence of the given series. Denote the moduli of z_0, z_1 by ρ_0, ρ_1, respectively. Then we must have $\rho_0 \leqq \rho_1$. If $\rho_0 = \rho_1$, we can take R equal to the common value. If $\rho_0 < \rho_1$, lay off upon the X-axis the distances ρ_0, ρ_1. Consider the point $a_1 = \dfrac{\rho_1 + \rho_0}{2}$. For $z = a_1$ the given series is either convergent or divergent. Let us suppose that it is convergent. Then by Theorem II the series is convergent for all values of z within the circle about the origin whose radius is a_1. The region of doubt lies now between the two circles of radii a_1, ρ_1, respectively. Consider the point $a_2 = \dfrac{\rho_1 + a_1}{2}$. For $z = a_2$ the series is again either convergent or divergent, say divergent. Then for values of z such that $|z| > a_2$ the series is divergent by Theorem III. The region of doubt now lies between the circles of radii a_1, a_2, respectively. Proceeding in this manner we shall obtain upon the X-axis an

infinite sequence of intervals each lying in the preceding one. Moreover, the length of the intervals has the limiting value zero. These intervals therefore define a definite number R. If we now describe a circle about the origin having R as a radius, we can say that the given power series converges for values of z for which $|z| < R$ and diverges for values of z for which $|z| > R$. For $z = R$ the series may or may not converge.

This circle whose radius is R is called the **circle of convergence** of the power series, and R is the **radius of convergence.** The radius of convergence may be equal to zero, in which case the given power series converges for $z = 0$ only, or it may be finite and different from zero, or it may be infinite, in which case the given series converges for all finite values of z. Nothing can be said from the discussion thus far concerning the convergence of the series for points on the circle of convergence itself. As a matter of fact a power series may converge absolutely at every point on the circle of convergence, or it may converge conditionally at every such point, or it may converge conditionally at certain points upon this circle and diverge at other points, or finally it may diverge at all points upon this circle.*

We shall need methods by which we may determine the radius of convergence of a given power series. It evidently depends upon the coefficients of the given series. A relation between the radius of convergence and the coefficients of the given series is given by the following theorem.

THEOREM IV. *If the coefficients of the given power series* $\Sigma \alpha_n z^n$ *are such that the limit* $\underset{n=\infty}{L} \left|\frac{\alpha_{n+1}}{\alpha_n}\right|$ *exists, then the value of this limit is equal to the reciprocal of R; that is, it is the reciprocal of the radius of convergence of the given series.*

Put

$$\underset{n=\infty}{L} \left|\frac{\alpha_{n+1}}{\alpha_n}\right| = A.$$

To prove that $\frac{1}{A}$ is equal to the radius of convergence, it is necessary to show that the given power series converges for all values of z where $|z| < \frac{1}{A}$ and diverges for all values of z where $|z| > \frac{1}{A}$.

* See *Encyclopédie des Sci. Math.*, II_7, p. 15.

We may readily show that the power series converges for values of z where $|z| < \frac{1}{A}$. As in the demonstration of Theorem I, let

$$\alpha_n = \rho_n(\cos\theta_n + i\sin\theta_n), \quad z = r(\cos\phi + i\sin\phi).$$

By Theorem III, Art. 42, the given power series converges if the series of moduli

$$\rho_0 + \rho_1 r + \rho_2 r^2 + \cdots + \rho_n r^n + \cdots \tag{1}$$

converges. This series converges if we have

$$\underset{n=\infty}{L} \frac{\rho_{n+1}r^{n+1}}{\rho_n r^n} = \underset{n=\infty}{L} \frac{\rho_{n+1}}{\rho_n}\cdot r < 1.$$

By the condition of the theorem we have

$$\underset{n=\infty}{L} \frac{\rho_{n+1}}{\rho_n} = \underset{n=\infty}{L} \left|\frac{\alpha_{n+1}}{\alpha_n}\right| = A.$$

Hence the condition that (1) converges is that

$$rA < 1;$$

that is,

$$|z| \equiv r < \frac{1}{A}.$$

Consequently, the given series converges for all values of z within the circle of radius $\frac{1}{A}$.

The given series likewise diverges for all values of z without the circle of radius $\frac{1}{A}$. To show this, suppose it to converge for some value z_0 without this circle. Let z_1 be any point outside of the circle such that $|z_1| < |z_0|$. Then by Theorem II the given series converges absolutely for z_1. We have then the convergent series

$$\rho_0 + \rho_1 r_1 + \cdots + \rho_n r_1^n + \cdots. \tag{2}$$

However, we have

$$\underset{n=\infty}{L} \frac{\rho_{n+1}r_1^{n+1}}{\rho_n r_1^n} = r_1 A > 1,$$

since $r_1 > \frac{1}{A}$. This result contradicts the conclusion that series (2) is convergent. From this contradiction it follows that $\Sigma\alpha_n z^n$ can not converge for any value of z exterior to the circle of radius $\frac{1}{A}$.

Since the given series converges for all values of z within the circle about the origin having the radius $\frac{1}{A}$ and is divergent for all values of z without this circle, it follows that $\frac{1}{A}$ must be equal to R, the radius of convergence, as the theorem states.

The application of Theorem IV to the problem of determining the radius of convergence depends upon the existence of the limit $\underset{n=\infty}{L}\left|\frac{\alpha_{n+1}}{\alpha_n}\right|$. The theorem gives a sufficient but not a necessary condition for convergence. There are convergent series for which this limit does not exist. The following series furnishes an illustration.

Ex. 2. Given the series

$$1+\frac{1}{2}z+\frac{1}{2\cdot 3}z^2+\frac{1}{2^2\cdot 3}z^3+\frac{1}{2^2\cdot 3^2}z^4+\cdots+\frac{1}{2^n\cdot 3^{n-1}}z^{2n-1}+\frac{1}{2^n\cdot 3^n}z^{2n}+\cdots.$$

This series is convergent for $|z|<1$; for putting $z=1$, we get a series whose terms are less than the corresponding terms of the convergent series $\sum\left(\frac{1}{2}\right)^n$. The limit $\underset{n=\infty}{L}\left|\frac{\alpha_{n+1}}{\alpha_n}\right|$ does not exist since $\left|\frac{\alpha_{n+1}}{\alpha_n}\right|$ oscillates between $\frac{1}{2}$ and $\frac{1}{3}$, depending upon whether an even or odd term is taken as the n^{th} term.

The following theorem * gives us a means of determining a radius of convergence that is applicable to any power series.

THEOREM V. *Given the series* $\sum_{n=0}^{\infty}\alpha_n z^n$; *and let* $\rho_n=|\alpha_n|$. *If* A *is the maximum limit of the sequence*

$$\rho_1,\ \sqrt{\rho_2},\ \sqrt[3]{\rho_3},\ \ldots,\ \sqrt[n]{\rho_n},\ \ldots, \tag{3}$$

then $\frac{1}{A}$ *is equal to the radius of convergence of the given series.*

(limit superior) - largest cluster pt.

By the maximum limit of a sequence is understood the largest number that can be obtained as the limit of a subsequence of the given sequence. In the particular case under discussion we are to consider the various subsequences that may be selected from (3) and denote by A the largest number that can be obtained as the limiting value of any of these subsequences.

* This theorem was first demonstrated by Cauchy. See his *Analyse Alg.*, p. 286, also *Encyclopédie des Sci. Math.*, II_7, p. 6.

We wish to show that the given series converges for

$$|z| < \frac{1}{A},$$

that is, within the circle C (Fig. 86), having the origin as a center and $R = \dfrac{1}{A}$ as a radius. Let z' be any point within the circle C. We have then

$$|z'| = \frac{1}{A+\epsilon}, \text{ where } 0 < \epsilon.$$

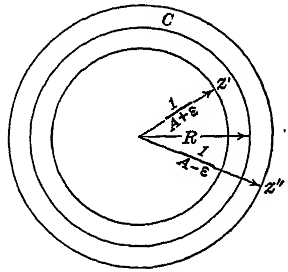

FIG. 86.

There are at most a finite number of elements of the sequence (3) greater than or equal to $A + \epsilon$. Suppose m is the largest of the subscripts of these elements. We have then

$$\frac{1}{|z'|} = A + \epsilon > \sqrt[n]{\rho_n}, \qquad n > m,$$

or

$$|z'|\sqrt[n]{\rho_n} < 1, \qquad n > m.$$

We therefore have

$$\rho_n |z'^n| = |\alpha_n z'^n| < 1, \qquad n > m.$$

It follows from Theorem I that the series $\sum\limits_{n=m+1}^{\infty} \alpha_n z^n$, and hence the given series, converges absolutely for all values of z within the circle whose radius is $\dfrac{1}{A+\epsilon}\cdot$ As ϵ is arbitrarily small it follows that the series $\Sigma \alpha_n z^n$ converges absolutely within the circle C.

We wish now to show that the given series diverges for

$$|z| > \frac{1}{A}.$$

Let z'' be any point exterior to the circle C. We have then

$$|z''| = \frac{1}{A-\epsilon}, \qquad \epsilon > 0.$$

There are now an infinite number of elements of the sequence (3) greater than $A - \epsilon$; that is, for an infinite number of values of n we have

$$\frac{1}{|z''|} = A - \epsilon < \sqrt[n]{\rho_n},$$

or

$$|z''|\ \sqrt[n]{\rho_n} > 1.$$

Then for $|z| \geqq |z''|$ we have, for an infinite number of values of n,

$$\rho_n\,|z^n| = |\alpha_n z^n| > 1.$$

Consequently, the given series $\Sigma\alpha_n z^n$ can not converge for $|z| > \frac{1}{A}$.

Since $\Sigma\alpha_n z^n$ is convergent for all values of z lying within the circle of radius $\frac{1}{A}$ and divergent for all values of z lying without this circle, it follows that $\frac{1}{A}$ is equal to R, the radius of convergence, which establishes the theorem.

Whenever the sequence (3) has a definite limit as n becomes infinite, the various subsequences have the same limit and hence the maximum limit is the limit of the sequence. It will be observed also that whenever both the sequence (3) and the ratio $\left|\frac{\alpha_{n+1}}{\alpha_n}\right|$ have a limit, the two limits are the same, since both are the reciprocal of the radius of convergence. Theorem V often enables us to determine the radius of convergence even if the sequence (3) has no definite limiting value.

Ex. 3. Determine the radius of convergence of the power series given in Ex. 2.

We have the sequence of positive values

$$\frac{1}{2},\ \sqrt{\frac{1}{2}\cdot\frac{1}{3}},\ \ldots,\ \sqrt[2n-1]{\left(\frac{1}{2}\right)^n\left(\frac{1}{3}\right)^{n-1}},\ \sqrt[2n]{\left(\frac{1}{2}\right)^n\left(\frac{1}{3}\right)^n},\ \ldots.$$

The limit of the subsequence in which the odd roots alone are taken is

$$\begin{aligned}\underset{n=\infty}{L}\sqrt[2n-1]{\left(\frac{1}{2}\right)^n\cdot\left(\frac{1}{3}\right)^{n-1}} &= \underset{n=\infty}{L}\left(\frac{1}{2}\right)^{\frac{n}{2n-1}}\cdot\left(\frac{1}{3}\right)^{\frac{n-1}{2n-1}}\\ &= \underset{n=\infty}{L}\left(\frac{1}{6}\right)^{\frac{n-1}{2n-1}}\underset{n=\infty}{L}\left(\frac{1}{2}\right)^{\frac{1}{2n-1}}\\ &= \frac{1}{\sqrt{6}}.\end{aligned}$$

The limit of the subsequence in which the even roots alone are taken is

$$\begin{aligned}\underset{n=\infty}{L}\sqrt[2n]{\left(\frac{1}{2}\right)^n\cdot\left(\frac{1}{3}\right)^n} &= \underset{n=\infty}{L}\left(\frac{1}{6}\right)^{\frac{n}{2n}}\\ &= \frac{1}{\sqrt{6}}.\end{aligned}$$

No other subsequence has a different limit and hence the sequence

$$\rho_1, \ \sqrt{\rho_2}, \ \sqrt[3]{\rho_3}, \ \ldots, \ \sqrt[n]{\rho_n}, \ \ldots$$

has the limit $\dfrac{1}{\sqrt{6}}$, and the radius of convergence of the given power series is $\sqrt{6}$.

The following theorem is of importance in the discussion of analytic functions.

Theorem VI. *The power series $\Sigma \alpha_n z^n$ converges uniformly in the closed region bounded by any circle about the origin whose radius is $R' < R$, where R is the radius of convergence of the given series. In the open region bounded by the circle of convergence the power series represents a function which is holomorphic.*

Let R'' be a y number such that

$$R' < R'' < R.$$

Then the series of positive terms

$$|\alpha_0| + |\alpha_1| R'' + |\alpha_2| R''^2 + \cdots \tag{4}$$

converges. For values of z such that $|z| \leqq R'$ the terms of the given series are less in absolute value than the corresponding terms of (4). Hence, by Theorem I, Art. 45, the given series converges absolutely and uniformly in the closed region bounded by the circle of radius R'.

Since any closed region bounded by an ordinary curve C and lying wholly in the open region bounded by the circle of convergence can be included within a circle of radius $R' < R$, it follows that the given power series is absolutely and uniformly convergent in every such closed region. Therefore, by Theorem II of the last article the given series represents a function holomorphic in the open region bounded by the circle of convergence, as stated in the theorem.

The foregoing theorem states nothing, however, as to the uniform convergence of the given power series in the open region bounded by the circle of convergence.

Ex. 4. Consider the convergence of the series

$$1 + \frac{z}{2} + \frac{z^2}{4} + \frac{z^3}{8} + \cdots + \frac{z^n}{2^n} + \cdots.$$

The circle of convergence has the radius

$$R = \frac{1}{\underset{n=\infty}{L} \sqrt[n]{\dfrac{1}{2^n}}} = 2.$$

The remainder $R_n(z)$ after the first n terms is

$$R_n(z) = \frac{z^n}{2^{n-1}(2-z)}.$$

As $|R_n(z)|$ for any value of n may be made as large as we choose by taking $|z|$ sufficiently near 2, it follows that there is no number m independent of z, such that, for all values of z within the circle of convergence,

$$|R_n(z)| < \epsilon, \qquad n > m.$$

Hence the series does not converge uniformly in the open region within the circle of convergence.

However, for all values of z within a circle about the origin having a radius $R' < 2$, there exists a number m such that for $n > m$, we have

$$|R_n(z)| \equiv \left|\frac{z^n}{2^{n-1}(2-z)}\right| < \epsilon,$$

Consequently in the closed region bounded by a circle of radius $R' < R$ the given series converges uniformly.

The following theorem gives a condition under which a power series is uniformly convergent in the closed region bounded by the circle of convergence.

THEOREM VII. *If $\Sigma\alpha_n z^n$ is absolutely convergent at one point on the circle of convergence, then it converges absolutely and uniformly in the closed region bounded by the circle of convergence.*

If the given series is absolutely convergent at one point on the circle of convergence, say at $z = z_0$, we know from the definition of absolute convergence that the series $\Sigma\,|\alpha_n|\,R^n$ converges, where R is the radius of convergence. However, any point on the circle whose radius is R gives the same series of moduli. Hence, for values of z such that $|z| \leqq R$ the terms of the given series are not greater in absolute value than the corresponding terms of $\Sigma\,|\alpha_n|\,R^n$. Hence by Theorem I, Art. 45, the given series converges absolutely and uniformly in the closed region bounded by the circle of radius R.

Ex. 5. Given the series $\sum \frac{z^{n^2}}{2^n}$, the character of whose convergence is to be examined.

This series is absolutely convergent for $z = 1$, since the series of moduli $\sum \frac{1}{2^n}$ is convergent. Hence, by Theorem VII the series converges absolutely and uniformly in the closed region bounded by the circle about the origin whose radius is 1.

The function represented by this series (Theorem II, Art. 45) is continuous in the closed region bounded by the unit circle, and by Theorem VI is holomorphic within this circle. This particular function,* however, is not holomorphic upon the unit circle itself.

COROLLARY. *If $\Sigma a_n z^n$ is divergent or conditionally convergent at any point on the circle of convergence, then it can be at best only conditionally convergent at any other point on this circle.*

This proposition follows as a consequence of Theorem VII; for, if the given series is absolutely convergent at any other point on the circle of convergence, then by Theorem VII, it must converge absolutely for all values of z for which $|z| \leqq R$, the radius of convergence, and this is a contradiction of the hypothesis. Hence, if the given series converges at any other points on the circle of convergence, it must converge conditionally.

Ex. 6. Consider the character of the series $\sum \frac{z^n}{n}$.

This series is conditionally convergent at $z = -1$. It is divergent at $z = 1$. Hence, in this case, the series can not be absolutely convergent at any point on the unit circle.

For the differentiation and integration of a power series we have the following theorem.

THEOREM VIII. *The power series $\Sigma a_n z^n$ may be differentiated or integrated term by term in the open region bounded by the circle of convergence. The circle of convergence of the resulting series is the same as that of the given series.*

That the given power series may be integrated or differentiated term by term in the open region bounded by the circle of convergence follows from Theorem II of the last article by the same reasoning as was employed in the demonstration of Theorem VI.

The resulting series in either case has the same circle of convergence as the original series. We shall show this to be true for term by term differentiation. A similar argument will establish the truth of the statement for term by term integration. The series of derivatives

$$f'(z) = u_1'(z) + u_2'(z) + \cdots + u_n'(z) + \cdots \tag{5}$$

is a power series. By the first part of the theorem under consideration the series (5) converges for all values of z within the circle of radius R. We must show that it is divergent for values of z exterior

* See Picard, *Traité d'analyse*, 2[d] Ed., Vol. 2, p. 74.

to the circle of radius R. Suppose it should converge for some value z_1 exterior to this circle. Then for values of z such that $|z| < R''$, where $R < R'' < |z_1|$, the series (5) converges uniformly and can be integrated term by term. As a result of integration we should have the original power series, except as to an additive constant, and this power series must then converge for all values of z such that $|z| < R''$. This, however, is impossible since values of z exterior to the circle of convergence of the given series are thus included. From this contradiction it follows that the series (5) can not converge for values of z exterior to the circle of radius R. Since the series (5) converges for all values of z within the circle of radius R and diverges for all values of z exterior to it, it follows that R is the radius of convergence of (5) as stated.

48. Expansion of a function in a power series. We have seen that in the open region bounded by the circle of convergence, a power series represents a function which is holomorphic. We shall now show that a function may be uniquely represented by a power series in the neighborhood of any point of a region in which it is holomorphic. Moreover, we shall develop a method for obtaining the required power series. The results may be stated in the following theorem.

THEOREM I. *If $f(z)$ is holomorphic in a given region S, then in the neighborhood of any point z_0 in S, $f(z)$ can be represented by a power series in $(z - z_0)$, and that in one and only one way, namely:*

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \frac{f''(z_0)}{2\,!}(z - z_0)^2 + \cdots + \frac{f^{[n]}(z_0)}{n!}(z - z_0)^n + \cdots.$$

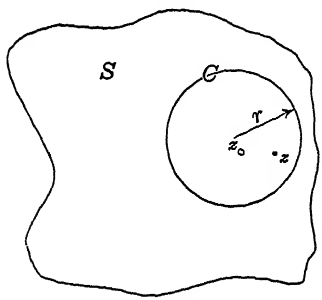

FIG. 87.

Let z_0 be any point in the given region S. About the point z_0 as a center draw the circle C, having the radius r and lying within S. Let the complex variable t take values corresponding to the points of C. For any point z within the circle, we have then

$$|z - z_0| < |t - z_0|, \quad \text{or} \quad \left|\frac{z - z_0}{t - z_0}\right| < 1.$$

Consider now the series

$$\frac{1}{t - z_0} + \frac{z - z_0}{(t - z_0)^2} + \frac{(z - z_0)^2}{(t - z_0)^3} + \cdots + \frac{(z - z_0)^n}{(t - z_0)^{n+1}} + \cdots. \quad (1)$$

This series converges for values of z within C and represents the function $\frac{1}{t-z}$; for, it is a geometric series having the ratio $\frac{z-z_0}{t-z_0}$. Considered as a series in the complex variable t, it converges uniformly upon the circle C; for, $|z-z_0|$ and $|t-z_0|=r$ are then both constant and the series of maximum numerical values

$$\frac{1}{r}+\frac{|z-z_0|}{r^2}+\frac{|z-z_0|^2}{r^3}+\cdots+\frac{|z-z_0|^n}{r^{n+1}}+\cdots$$

converges. The uniformity of the convergence is not disturbed if we multiply each term by $f(t)$. We thus obtain

$$\frac{f(t)}{t-z}=\frac{f(t)}{t-z_0}+\frac{(z-z_0)f(t)}{(t-z_0)^2}+\frac{(z-z_0)^2f(t)}{(t-z_0)^3} \\ +\cdots+\frac{(z-z_0)^nf(t)}{(t-z_0)^{n+1}}+\cdots. \tag{2}$$

Since this series converges uniformly, we may integrate it term by term with respect to t, the integral being taken around the circle C. We thus obtain

$$f(z)=\frac{1}{2\pi i}\int_C\frac{f(t)\,dt}{t-z}=\frac{1}{2\pi i}\int_C\frac{f(t)\,dt}{t-z_0}+(z-z_0)\frac{1}{2\pi i}\int_C\frac{f(t)\,dt}{(t-z_0)^2} \\ +\cdots+(z-z_0)^n\frac{1}{2\pi i}\int_C\frac{f(t)\,dt}{(t-z_0)^{n+1}}+\cdots. \tag{3}$$

Each of these integrals is a constant, and we have by Art. 20,

$$\frac{f^{(n)}(z_0)}{n!}=\frac{1}{2\pi i}\int_C\frac{f(t)\,dt}{(t-z_0)^{n+1}}, \qquad n=0,1,2,\ldots.$$

Replacing the integrals in (3) by their equivalent values in terms of the successive derivatives of $f(z)$, we have for $z=z_0$ the required expansion known as **Taylor's series**, namely:

$$f(z)=f(z_0)+f'(z_0)(z-z_0)+\frac{f''(z_0)}{2!}(z-z_0)^2 \\ +\cdots+\frac{f^{(n)}(z_0)}{n!}(z-z_0)^n+\cdots. \tag{4}$$

For $z_0=0$ we have for the expansion of the given function in the neighborhood of the origin a special form of Taylor's series known as **Maclaurin's series**, namely:

$$f(z)=f(0)+f'(0)z+\frac{f''(0)}{2!}z^2+\cdots+\frac{f^{(n)}(0)}{n!}z^n+\cdots. \tag{4'}$$

Within the circle C, that is in a neighborhood of z_0, the given function $f(z)$ can therefore be represented by a power series. Moreover, within C the given function can be represented by a power series in $(z - z_0)$ in only one way. For convenience put

$$\frac{f^{(n)}(z_0)}{n!} = \alpha_n.$$

The series (4) can then be written

$$\alpha_0 + \alpha_1 (z - z_0) + \alpha_2(z - z_0)^2 + \cdots + \alpha_n(z - z_0)^n + \cdots. \tag{5}$$

Suppose it is possible that within a circle C_1, whose radius is equal to or less than that of C, $f(z)$ can be represented by a second power series, say

$$f(z) = \beta_0 + \beta_1(z - z_0) + \beta_2(z - z_0)^2 + \cdots + \beta_n(z - z_0)^n + \cdots. \tag{6}$$

Subtracting (6) from (5) we have

$$0 = (\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)(z - z_0) + (\alpha_2 - \beta_2)(z - z_0)^2 + \cdots + (\alpha_n - \beta_n)(z - z_0)^n + \cdots. \tag{7}$$

This relation holds for all values of z within C_1, hence for $z = z_0$. Putting $z = z_0$, we get

$$0 = \alpha_0 - \beta_0, \quad \text{or} \quad \alpha_0 = \beta_0.$$

For $z \neq z_0$, however, we have

$$0 = (\alpha_1 - \beta_1)(z - z_0) + (\alpha_2 - \beta_2)(z - z_0)^2 + \cdots + (\alpha_n - \beta_n)(z - z_0)^n + \cdots.$$

Dividing by $(z - z_0)$, we obtain

$$0 = (\alpha_1 - \beta_1) + (\alpha_2 - \beta_2)(z - z_0) + \cdots + (\alpha_n - \beta_n)(z - z_0)^{n-1} + \cdots.$$

This series converges uniformly within or upon any circle about z_0 as a center and lying within C_1. Hence it defines a continuous function, say $\phi(z)$. Since $\phi(z)$ is continuous and equal to zero for $z \neq z_0$, we have for $z = z_0$,

$$\phi(z_0) = \underset{z \doteq z_0}{L}\, \phi(z) = \underset{z \doteq z_0}{L}\, 0 = 0.$$

Consequently, we have

$$0 = \phi(z_0) = \alpha_1 - \beta_1 = 0, \quad \text{or} \quad \alpha_1 = \beta_1.$$

Continuing in this manner we may show that

$$\alpha_n = \beta_n, \qquad n = 0, 1, 2, \ldots;$$

hence, in the neighborhood of z_0 the given function can be represented by a power series in $(z - z_0)$ in only one way. Since z_0 is any point in the given region S, the theorem is established.

Ex. Expand $f(z) = \dfrac{z}{1-z}$ in a Maclaurin's series.

We have

$$\begin{aligned} f(o) &= 0, & & \\ f'(z) &= (1-z)^{-2}, & f'(0) &= 1, \\ f''(z) &= 2\,(1-z)^{-3}, & \frac{f''(0)}{2\,!} &= 1, \\ f'''(z) &= 2\cdot 3\,(1-z)^{-4}, & \frac{f'''(0)}{3\,!} &= 1, \\ f^{(n)}(z) &= n\,!\,(1-z)^{-(n+1)}, & \frac{f^{(n)}(0)}{n\,!} &= 1. \end{aligned}$$

Hence, in the neighborhood of the origin the series

$$z + z^2 + z^3 + \cdots + z^n + \cdots \tag{8}$$

represents the given function.

It is to be noted that the power series in $(z - z_0)$ arising by Taylor's expansion of a given function which is holomorphic in a region S represents that function for all values of z within any circle that can be drawn about the given point z_0, so long as it lies within the given region S and incloses only points of S; for, it is clear that any such circle can be selected as the curve C along which the integrals are taken that determine the coefficients in the expansion. Moreover, if as in the illustration given above, the function is holomorphic within the entire circle of convergence, the series represents the function giving rise to it within the whole of that circle.

We now have shown that every power series defines a function which is holomorphic in the open region bounded by the circle of convergence, and conversely that a function can be expanded in a power series in the neighborhood of any point in the region S in which it is holomorphic. It will be seen, therefore, that power series play an important rôle in the discussion of analytic functions. Indeed, Weierstrass based his entire development of the theory of analytic functions upon power series.

EXERCISES

1. Determine the circle of convergence of the series

(a) $1 + 4z + 9z^2 + 16z^3 + \cdots,$

(b) $1 + mz + \dfrac{m(m-1)}{2\,!}z^2 + \cdots,$

(c) $1 + z + \dfrac{z^2}{2\,!} + \dfrac{z^3}{3\,!} + \cdots,$

(d) $z - \dfrac{z^2}{2} + \dfrac{z^3}{3} - \dfrac{z^4}{4} + \cdots,$

(e) $1 + \dfrac{z}{\lambda} + \dfrac{z^2}{\lambda^2} + \dfrac{z^3}{\lambda^3} + \cdots.$

2. Discuss the uniform convergence of the series

(a) $1 + \dfrac{z}{\lambda} + \dfrac{z^2}{\lambda^2} + \dfrac{z^3}{\lambda^3} + \cdots + \dfrac{z^n}{\lambda^n} + \cdots,$

(b) $1 + \dfrac{z}{\lambda} + \dfrac{z^4}{\lambda^2} + \dfrac{z^9}{\lambda^3} + \cdots + \dfrac{z^{n^2}}{\lambda^n} + \cdots,$

where $|\lambda| > 1$. What can be said of the function represented by the second series that can not be said of the function represented by the first?

3. Discuss the behavior of the series

$$1 + mz + \frac{m(m-1)}{2\,!}z^2 + \frac{m(m-1)(m-2)}{3\,!}z^3 + \cdots$$

for values of z upon the circle of convergence,*

(a) for $m > 0$,

(b) for $m \leqq -1$,

(c) for $-1 < m \leqq 0$.

4. Show that the two series

(a) $\dfrac{1}{2}\,\dfrac{1-z}{1} + \dfrac{1\cdot 3}{2\cdot 4}\,\dfrac{(1-z)^2}{2} + \dfrac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\,\dfrac{(1-z)^3}{3} + \cdots,$

(b) $\dfrac{1}{2}\,\dfrac{z-1}{1} + \dfrac{1\cdot 3}{2\cdot 4}\,\dfrac{(z-1)^2}{2} + \dfrac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\,\dfrac{(z-1)^3}{3} + \cdots$

have the same region of convergence.

5. Determine the region of convergence of the series

$$\frac{z-1}{z} + \frac{1}{2}\left(\frac{z-1}{z}\right)^2 + \frac{1}{3}\left(\frac{z-1}{z}\right)^3 + \cdots.$$

Find the derivative of the function represented by the given series.

6. Given the series

$$\frac{1}{1\cdot 2\cdot 3} + \frac{z}{2\cdot 3\cdot 4} + \frac{z^2}{3\cdot 4\cdot 5} + \cdots.$$

Verify, by testing the first and second derived series, that they have the same circle of convergence as the original series.

* See Goursat, *Cours d'analyse mathematique*, 2d Ed. (1911), Vol. 2, p. 43.

7. Given the series

$$\cos z - \frac{\cos^3 z}{3} + \frac{\cos^5 z}{5} - \cdots + \frac{(-1)^{n-1}\cos^{2n-1} z}{2n-1} + \cdots.$$

For what values of $\cos z$ is the series convergent? Determine the corresponding region in the Z-plane. Does the series represent a continuous function of z in this region?

8. Given the series

$$z^2 + \frac{z^2}{1+z^2} + \frac{z^2}{(1+z^2)^2} + \cdots.$$

Show that this series converges for all finite values of z outside the lemniscate

$$\rho^2 = -2\cos 2\theta.$$

Show that this series diverges at all points different from zero within and upon this lemniscate.

9. Derive the following expansions and determine in each case the region of convergence:

(a) $$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots,$$

(b) $$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots,$$

(c) $$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots,$$

(d) $$\log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \cdots,$$

(e) $$(1+z)^m = 1 + mz + \frac{m(m-1)}{2!}z^2 + \cdots.$$

10. Derive the expansion

$$\frac{1}{1+z} = 1 - z + z^2 - \cdots$$

and verify the expansion in Ex. 9 (d) by integration of this series. Derive in a similar manner the expansion

$$\arctan z = \int_0^z \frac{dz}{1+z^2} = z - \frac{z^3}{3} + \frac{z^5}{5} - \cdots.$$

11. Making use of the expansions in Ex. 9, derive the following relations:

(a) $$\cos 2z = \cos^2 z - \sin^2 z,$$

(b) $$\log\left(\frac{1}{1-z}\right) = \log 1 - \log(1-z),$$

(c) $$\tan z = z + \frac{1}{3}z^3 + \frac{2}{15}z^5 + \cdots,$$

(d) $$\sinh z = \frac{e^z - e^{-z}}{2} = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \cdots,$$

(e) $$\cosh z = \frac{e^z + e^{-z}}{2} = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \cdots.$$

Verify these results by Taylor's theorem.

12. From the expansion

$$\tan z = z + \frac{1}{3} z^3 + \frac{2}{15} z^5 + \cdots$$

derive the expansion

$$\cot z = \frac{1}{z} - \frac{1}{3} z - \frac{1}{45} z^3 + \cdots.$$

13. Making use of the expansions

$$\sin z = z - \frac{z^3}{3\,!} + \frac{z^5}{5\,!} - \frac{z^7}{7\,!} + \cdots,$$

$$\cos z = 1 - \frac{z^2}{2\,!} + \frac{z^4}{4\,!} - \frac{z^6}{6\,!} + \cdots,$$

derive the expansions

$$(a)\quad \csc z = \frac{1}{z} + \frac{z}{3\,!} + \frac{7 z^3}{3 \cdot 5\,!} + \cdots,$$

$$(b)\quad \sec z = 1 + \frac{z^2}{2\,!} + \frac{5 z^4}{4\,!} + \cdots.$$

14. Derive the expansions

$$(a)\quad \operatorname{arc}\sin z = \int_0^z \frac{dz}{\sqrt{1 - z^2}} = z + \frac{1}{2}\frac{z^3}{3} + \frac{1 \cdot 3}{2 \cdot 4}\frac{z^5}{5} + \cdots,$$

$$(b)\quad \int \frac{z\,dz}{\cos z} = \frac{z^2}{2} + \frac{z^4}{4 \cdot 2\,!} + \frac{5 z^6}{6 \cdot 4\,!} + \cdots.$$

15. Verify the formulæ

$$\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta,$$
$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$

by means of the power series expansion of the sine and cosine.

16. Given the expansion

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots.$$

By aid of this series prove that

$$e^{z_1} \cdot e^{z_2} = e^{z_1 + z_2}$$

and give the reason for each step.

CHAPTER VII

GENERAL PROPERTIES OF SINGLE-VALUED FUNCTIONS

49. Analytic continuation. In the present chapter we shall discuss some of the general properties of single-valued functions. Let us first consider how a function which is holomorphic in a certain region may be completely represented in that region by means of power series.

Consider, for example, the function

$$f(z) = \frac{1}{1 - z}.$$

Expanding this function in a Maclaurin series, we have

$$1 + z + z^2 + z^3 + \cdots + z^n + \cdots. \tag{1}$$

This series converges within the circle C of unit radius about the origin. Since the given function $f(z)$ is holomorphic for all finite values of z, except $z = 1$, it follows from Art. 48 that the series (1) represents that function for all values of z within C. However, for values of z exterior to C the series (1) does not converge and hence can not be said to represent the given function. We may for convenience denote the aggregate of functional values corresponding to values of z within C, as given by the series (1), by the symbol $\phi(z)$. We shall speak of $\phi(z)$ as an **element** of the function $f(z) = \dfrac{1}{1 - z}$. A general definition of an element of an analytic function will be given later in this article. Since $f(z)$ can be represented by a Taylor's expansion in the neighborhood of any point of a region in which the function is holomorphic, there is similarly an element $\phi_0(z)$ corresponding to any finite point z_0 of the complex plane, except the point $z = 1$. The power series defining these respective elements of $f(z)$ converge within circles which may overlap.

For example let z_0 be a point within C, so selected that for some values of z upon C we have $|z - z_0| < \dfrac{|1 - z_0|}{2}$. Expanding the given function $\dfrac{1}{1 - z}$ in powers of $(z - z_0)$, that is in a Taylor's series, we

get

$$\frac{1}{1-z_0}+\frac{z-z_0}{(1-z_0)^2}+\frac{(z-z_0)^2}{(1-z_0)^3}+\cdots+\frac{(z-z_0)^{n-1}}{(1-z_0)^n}+\cdots. \qquad (2)$$

This series is a geometric series having the ratio $\frac{z-z_0}{1-z_0}$, and hence it converges for values of z such that

$$|z-z_0|<|1-z_0|;$$

that is, it converges for values of z within a circle of C_0 of radius $|1-z_0|$ about z_0 as a center. Since the point z_0 was so selected that at least one point on C is closer to z_0 than one-half of the distance $|1-z_0|$, it follows that C_0 must intersect C.

In that portion of the plane included within these two circles of convergence, the given function $\frac{1}{1-z}$ is represented by either of the two series, each giving the same numerical value for any particular value of z within the common region. Consider now an assemblage of power series obtained from $\frac{1}{1-z}$, such that the corresponding circles of convergence cover the entire finite portion of the plane except the one point $z=1$, which is not a regular point of the given function. This assemblage of power series may be said to completely represent the given function.

In the foregoing illustration the function $f(z)$ is given by means of an algebraic expression in z, from which the value of the function can be computed for any value of z except $z=1$. From this expression we are able to obtain an expansion of the function in a power series in the neighborhood of any point of the region in which the function is holomorphic. We shall now show that had we known merely the values of the function in ever so small a neighborhood of any point of the complex plane other than $z=1$, say the origin, we should have been able, at least theoretically, to compute the value of the function at every point of the region in which it is holomorphic and that without even finding the expression $\frac{1}{1-z}$ at all. The unique determination of the values of a function in a more or less extended region by its values in an arbitrarily small portion of that region is a general property of functions of a complex variable for regions in which they are holomorphic. This property may be more exactly formulated in the following theorem.

THEOREM I. *If a function $f(z)$ is holomorphic in a given region S, then it is uniquely determined for all values of z in S by its values along any arbitrarily small arc of an ordinary curve proceeding from a point of S.*

Let α be a point of the given region S from which the given arc is drawn and suppose β to be any other point of S. Let α and β be connected by any ordinary curve $\mathcal{L}$ lying wholly within S and coinciding with the given arc in the neighborhood of α. Let $f(z)$ be holomorphic in S and suppose its values to be given along that portion of $\mathcal{L}$ lying in an arbitrarily small neighborhood of α. We are to show that $f(\beta)$ is then uniquely determined. Since $f(z)$ is holomorphic in the neighborhood of α, its derived functions are also holomorphic in the same neighborhood, and hence for $z = \alpha$ the successive derivatives

$$f'(\alpha),\ f''(\alpha),\ \ldots,\ f^{(n)}(\alpha),\ \ldots \tag{3}$$

all exist. The existence of the derivative $f'(\alpha)$ implies that the limit

$$\underset{\Delta z=0}{L} \frac{f(\alpha+\Delta z)-f(\alpha)}{\Delta z},$$

is the same, when Δz approaches zero in any manner whatever. Hence, the value of $f'(\alpha)$ may be found from the given values of $f(z)$ by taking this limit as z approaches α along any curve proceeding from α, for example along the given curve $\mathcal{L}$. The higher derivatives are likewise determined by the given values of $f(z)$. Knowing the value of $f(z)$ for $z = \alpha$ and the successive derivatives given in (3) we may now write out Taylor's expansion of $f(z)$ for values of z in the neighborhood of α, namely,

$$f(\alpha)+f'(\alpha)(z-\alpha)+\frac{f''(\alpha)}{2!}(z-\alpha)^2+\cdots+\frac{f^{(n)}(\alpha)}{n!}(z-\alpha)^n+\cdots. \tag{4}$$

This series converges and defines an element $\phi_0(z)$ which is equal to $f(z)$ for all values of z within any circle drawn about α as a center and lying wholly within the region S. Let C_0 be a circle satisfying these conditions. If the point β lies within C_0, then the value of $f(\beta)$ is already seen to be uniquely determined; for, in order to find this value all we need to do is to substitute β for z in series (4). If β lies outside of C_0, let α_1 be a point of intersection of the curve $\mathcal{L}$ with C_0. Take a point z_1 on the given curve arbitrarily close to α_1 but within C_0. The function $f(z)$ is holomorphic in the neighborhood of $z = z_1$ and the successive derivatives of $f(z)$ for this value of z can be found by successively differentiating (4) term by term and

substituting z_1 for z in the several derived series. The coefficients of Taylor's expansion of $f(z)$ about the point z_1 are therefore uniquely determined. The resulting expansion is

$$f(z_1)+f'(z_1)(z-z_1)+\frac{f''(z_1)}{2!}(z-z_1)^2+\cdots+\frac{f^{(n)}(z_1)}{n!}(z-z_1)^n+\cdots. \quad (5)$$

This series in turn defines an element $\phi_1(z)$ which is identical with $f(z)$ for all values of z within a circle C_1 drawn about z_1 as a center and lying wholly in S. Since z_1 was taken arbitrarily close to α_1 and since α_1 is an inner point of S, the circle C_1 must intersect C_0. If β lies within the circle C_1, the value of $f(\beta)$ is uniquely determined; for, to find $f(\beta)$ we need merely to replace z by β in (5). If β lies outside of C_1, then take a point z_2 on $\mathcal{L}$ lying arbitrarily near the point α_2 where $\mathcal{L}$ cuts C_1 and compute as before the Taylor's expansion of $f(z)$ for values of z in the neighborhood of the point $z = z_2$. Proceeding in this manner, it is possible, at least theoretically, to obtain after a finite number of operations a series which converges within a circle C_k lying wholly within S and having β as an inner point. By substituting β for z in this series the value of $f(\beta)$ can be found and hence the given function is uniquely determined for $z = \beta$. However, β is any point of S and hence the given function $f(z)$ is uniquely determined for all points of S as stated in the theorem.

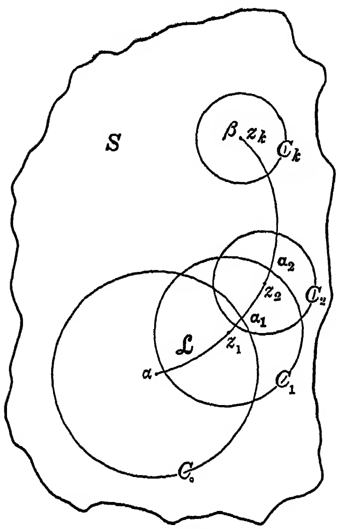

FIG. 88.

As a direct consequence of the foregoing theorem, we have the following corollary.

COROLLARY. *If two functions are holomorphic in a given region S and are equal for all values of z in the neighborhood of a point $z = \alpha$ of S, or for all values of z along an arbitrarily small arc proceeding from α, then the two functions are equal for all values of z in S.*

Thus far we have discussed for the most part functions which were known to be holomorphic in a given region. We have not inquired into the question as to how large that region might be in any

given case. By aid of the foregoing corollary we can now consider the possibility of extending the region in which a function is known to be holomorphic. Moreover, we shall be able to show that an analytic function is completely and uniquely determined if it is known for any region however small that region may be.

Let $\phi_1(z)$ be defined for the region S_1, Fig. 89, and in this region

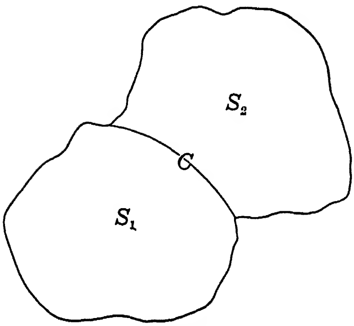

FIG. 89.

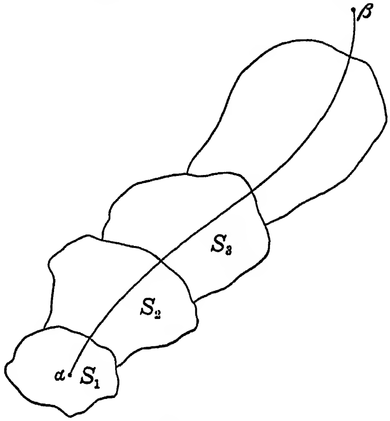

FIG. 90.

let it be holomorphic. Suppose it is possible to find a second function $\phi_2(z)$ which is holomorphic in an adjacent region S_2, having an arc C of an ordinary curve for at least a portion of the boundary between S_1 and S_2. Moreover, let $\phi_1(z)$, $\phi_2(z)$ be each defined for values of z along C, end points excepted, and for these values let each of these functions be holomorphic and equal to the other. Then for values of z in S_1, S_2 and along C the functions $\phi_1(z)$, $\phi_2(z)$ define a function $f(z)$ which is holomorphic in this enlarged region. The function $\phi_2(z)$ is called an **analytic continuation** of $\phi_1(z)$, and the process of finding such a function is called the process of analytic continuation.

It follows from the corollary to Theorem I that for values of z in S_2, the function $f(z)$ thus defined is uniquely determined. For suppose that another analytic continuation of $\phi_1(z)$, say $\Phi_1(z)$, could be found such that in the region S_2 it has values different from $\phi_2(z)$. We should have a function $F(z)$, defined by $\phi_1(z)$ and $\Phi_1(z)$, which is also holomorphic in the region $S_1 + S_2 + C$, that is, defined for values of z in S_1, S_2 and along the arc C. We then have two functions $f(z)$ and $F(z)$ each holomorphic in the region $S \equiv S_1 + S_2 + C$ and

equal to each other in S_1. By the foregoing corollary these two functions must be identical throughout the region S.

It is evident that $\phi_1(z)$ likewise may be regarded as an analytic continuation of $\phi_2(z)$. Either of these functions is uniquely determined when the other is known. Instead of having merely boundary points in common, the two regions S_1, S_2 may of course overlap. At the points common to the regions S_1, S_2 the two functions $\phi_1(z)$, $\phi_2(z)$ must in this case satisfy the same conditions as at points along the arc C in the case where the regions are adjacent but do not overlap, namely, they must have equal values and be holomorphic. In both cases we speak of the functions $\phi_1(z)$, $\phi_2(z)$ as **elements of the function** $f(z)$.

If an element $\phi_1(z)$ of a function is holomorphic in the neighborhood of a point, we say that $\phi_1(z)$ is continued analytically along a curve from α to β if this curve lies wholly within a finite sequence of connected regions $S_1, S_2, \ldots, S_n$ such that $\phi_1(z)$ has an analytic continuation $\phi_2(z)$ defined for S_2 and $\phi_2(z)$ likewise can be continued analytically in the region S_3, etc.

The following theorem * sets forth the necessary and sufficient conditions for analytic continuation.

THEOREM II. *Given two functions* $\phi_1(z)$, $\phi_2(z)$ *which are holomorphic respectively in the adjacent regions* S_1, S_2 *having an arc* C *of an ordinary curve as that portion of their boundaries common to the two. The necessary and sufficient condition that each of these functions is an analytic continuation of the other is that they converge uniformly to equal values on* C.

It follows at once from the definition of analytic continuation that the conditions set forth in the theorem are necessary; for, if either is an analytic continuation of the other, then they must have equal values along C, and moreover, for these values they must be holomorphic and hence continuous with the values at points within S_1, S_2, respectively.

To show that the given conditions are sufficient we define $f(z)$ as follows:

$$\begin{aligned} f(z) &= \phi_1(z) \text{ in } S_1 \\ &= \phi_2(z) \text{ in } S_2 \\ &= \phi_1(z) = \phi_2(z) \text{ along } C. \end{aligned}$$

* See Painlevé, *Toulouse Annales*, Vol. II, p. 28; also Pompeiu, *L'Enseignement Mathématique*, July, 1913, p. 305.

We must then show that $f(z)$ is holomorphic in the region S composed of S_1, S_2 and the points of C, end points excepted.

Let γ_1, γ_2 be two ordinary curves joining any two points A, B of C and lying wholly within the regions S_1, S_2, respectively. By the Cauchy-Goursat theorem, we have

$$\int_{\gamma_1} \phi_1(z)\, dz + \int_{AB} \phi_1(z)\, dz = 0, \qquad (6)$$

$$\int_{\gamma_2} \phi_2(z)\, dz + \int_{BA} \phi_2(z)\, dz = 0. \qquad (7)$$

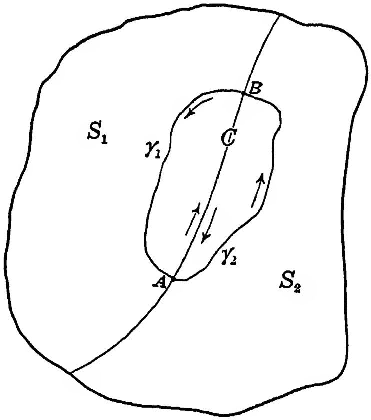

FIG. 91.

Denote by γ the curve $\gamma_1 + \gamma_2$. Combine the integrals in (6) with the corresponding integrals of (7). Since $\phi_1(z) = \phi_2(z)$ for values of z along C, we have

$$\int_{AB} \phi_1(z)\, dz + \int_{BA} \phi_2(z)\, dz = 0.$$

By definition $f(z)$ is equal to $\phi_1(z)$ for values of z in S_1 and to $\phi_2(z)$ for values of z in S_2. Hence, we have from (6) and (7)

$$\int_{\gamma} f(z)\, dz = \int_{\gamma_1} \phi_1(z)\, dz + \int_{\gamma_2} \phi_2(z)\, dz = 0. \qquad (8)$$

Since γ is any ordinary curve lying wholly within S and passing through A, B, it follows from Morera's theorem that $f(z)$ is holomorphic for all values of z in S and hence for all values of the variable along the curve C, end points excepted. Consequently, either of the two functions ϕ_1, ϕ_2 may be regarded as the analytic continuation of the other. Moreover, it follows from our previous discussion that either of these functions is uniquely determined when the other is known.

One of the methods of analytic continuation most frequently employed in theoretical discussions is that by means of power series. Suppose we have given a power series, say of the form

$$\alpha_0 + \alpha_1 (z - z_0) + \alpha_2 (z - z_0)^2 + \cdots + \alpha_n (z - z_0)^n + \cdots. \qquad (9)$$

Within its circle of convergence C_0 this series defines an element $\phi_0(z)$ which is holomorphic within C_0. Let z_1 be any point within C_0 but lying arbitrarily close to C_0. Since $\phi_0(z)$ is holomorphic in the

neighborhood of $z = z_1$, we can compute Taylor's expansion of $\phi_0(z)$, obtaining

$$\phi_0(z_1) + \phi_0'(z_1)(z - z_1) + \frac{\phi_0''(z_1)}{2!}(z - z_1)^2 + \cdots + \frac{\phi_0^{(n)}(z_1)}{n!}(z - z_1)^n + \cdots. \tag{10}$$

Let C_1 be the circle of convergence of the series (10); then within C_1 this series defines an element $\phi_1(z)$ which is holomorphic within C_1. Suppose C_1 intersects C_0; then in the region included within the two circles C_0 and C_1, the elements $\phi_0(z)$, $\phi_1(z)$ satisfy the conditions of analytic continuation and consequently define a function $f(z)$ which is holomorphic in this region. Either of these elements may be again continued by the same means. By this process any given element may, at least theoretically, be continued analytically by the method of power series until the whole complex plane, with the exception of certain points or regions excluded by the inherent character of the function $f(z)$ so defined, is covered by overlapping circles of convergence. The character of the exceptional points and regions will be discussed in subsequent articles.

The importance of the power series method of analytic continuation is due largely to its theoretical value. Other methods, although restricted in their uses for theoretical discussions, are of much greater practical importance in the applications of the theory of functions. We shall now consider a method introduced by Schwarz,* in which he makes use of the principle of symmetry. Let $\phi_1(z)$ be a function which is holomorphic in a region S_1 lying in the upper half-plane and having a segment AB of the axis of reals as a part of its boundary. Suppose that as z approaches any point x of AB along any path whatsoever lying interior to S_1, $\phi_1(z)$ approaches a definite real value $\phi_1(x)$. Then by Art. 13 $\phi_1(x)$ is a continuous function of x. Denote by $\bar{z}$ a point in the lower half-plane situated symmetrically with respect to z relative to the axis of reals. The assemblage of points $\bar{z}$ constitutes a region S_2 symmetrical to S_1 with respect to AB. Associate with each value of $\bar{z}$ a functional value which is the conjugate imaginary of $\phi_1(z)$. The assemblage of these values defines a function $\phi_2(\bar{z})$ which is holomorphic in S_2 and converges to the real values $\phi_2(x) = \phi_1(x)$ along the axis of reals.

* See *Crelle*, Vol. LXX, pp. 106, 107; also *Mathematische Abhandlungen*, Vol. II, pp. 65–83.

In the continuous region S made up of S_1, S_2 and the points along the axis of reals between A and B, the functions $\phi_1(z)$, $\phi_2(z)$ satisfy the conditions of Theorem II and hence $\phi_2(z)$ is an analytic continuation of $\phi_1(z)$. Each of these functions is then an element of a function $f(z)$ which is holomorphic in S and is equal to $\phi_1(z)$ in S_1 and equal to $\phi_2(z)$ in S_2, and moreover $f(z)$ takes the common values of the elements $\phi_1(z)$, $\phi_2(z)$ along the axis of reals between A and B. The advantage of this method of analytic continuation is the ease with which the continuation can be obtained. All that is needed is to reflect the given region upon the X-axis and associate with the reflected region a function which is the conjugate imaginary function of $\phi_1(z)$.

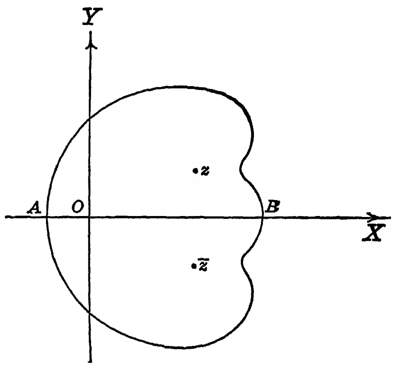

FIG. 92.

We shall now consider a generalization of the foregoing method of analytic continuation. To do so we shall make use of a generalization of the idea of reflection. Let the points of the segment AB of the axis of reals, which formed a common portion of the boundary between the two given regions, be made to correspond to the points of a regular arc C of an analytic curve. By an **analytic curve** is understood one whose parametric equations are of the form

$$x = \Psi_1(t), \qquad y = \Psi_2(t), \tag{11}$$

where $\Psi_1(t)$, $\Psi_2(t)$ are real, analytic functions of the real variable t. An arc of such a curve is regular if we have the added condition that the derivatives $\Psi_1'(t)$, $\Psi_2'(t)$ are not simultaneously zero; that is, if we have

$$[\Psi_1'(t)]^2 + [\Psi_2'(t)]^2 \neq 0, \qquad t_A < t < t_B.$$

To any point t_0 of AB there corresponds a point $z_0 \equiv (x_0, y_0)$ of C. As the two functions $\Psi_1(t)$, $\Psi_2(t)$ are analytic, each may be expanded in powers of $(t - t_0)$. The resulting series converge for all values of the variable within their circles of convergence, and hence t may take complex as well as real values. Denoting these real and complex values by τ, we have

$$z = x + iy = \Psi_1(\tau) + i\Psi_2(\tau) = \Psi(\tau), \tag{12}$$

which is holomorphic and has a derivative different from zero for all points of AB, end points at most excepted. The function $z = \Psi(\tau)$ is then defined for a region S of the τ-plane consisting of the inner points of AB and certain regions S_1, S_2 lying symmetrically with respect to AB, Fig. 93. By Theorem II, Art. 21, there exists in the Z-plane a corresponding region S' consisting of the points of C and the regions S_1', S_2' lying on either side of C, in which the inverse function $\tau = \phi(z)$ is uniquely determined and holomorphic. The region S can be so restricted that the function $z = \Psi(\tau)$ and its inverse function $\tau = \phi(z)$ map the regions S, S' upon each other.

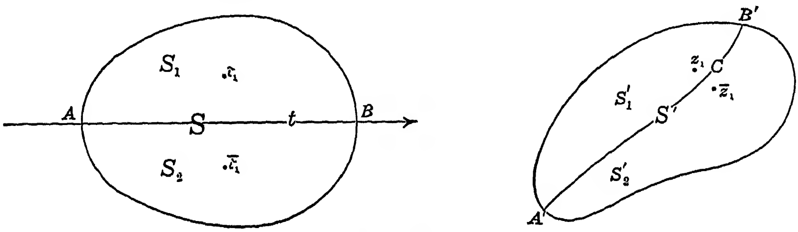

FIG. 93.

Then to any conjugate imaginary points τ_1 and $\bar{\tau}_1$ lying respectively in S_1 and S_2, there are associated two corresponding z-points namely z_1 and $\bar{z}_1$ lying respectively in S_1' and S_2', and conversely. It is to be noted that the particular values of z thus associated depend upon the form of the curve C and not upon the form of the parametric equations (11) of the curve. Suppose, for example, a different parametric representation of the curve C is obtained by replacing t in (11) by an analytic function of any other real variable r. If we then permit r to take complex values, conjugate imaginary points in the τ-plane correspond to conjugate imaginary points in the r-plane, and consequently we get the same corresponding values of z.

Of the two z-points corresponding to conjugate imaginary values of a parameter τ, either is said to be the **reflection** or image of the other with respect to the curve C. Likewise the region S_2' may be spoken of as the reflection of the region S_1' with respect to C.

This definition of reflection with respect to a regular arc of an analytic curve may now be used in developing a method of analytic continuation. Let S_1', S_2' be any two adjacent regions such that S_2' is the reflection of S_1' with respect to the regular arc C of an analytic curve whose parametric equations are $x = \Psi_1(t)$, $y = \Psi_2(t)$. Let $\phi_1(z)$ be a function which is holomorphic in S_1' and defined for values

of z along C by its limiting values as z approaches the points of C by any path whatever lying wholly within S_1'. We can now state in the following form the necessary and sufficient condition that $\phi_1(z)$ may be analytically continued by reflection with respect to arc C.

THEOREM III. *The necessary and sufficient condition that $\phi_1(z)$ can be analytically continued by reflection with respect to the regular arc C of an analytic curve forming a portion of the boundary of the region for which $\phi_1(z)$ is defined is that $\phi_1(z)$ converges uniformly to real values along C.*

If $\phi_1(z)$ can be analytically continued across the arc C into the region S_2', which is a reflection of S_1' with respect to C, then for the region S_2' a function $\phi_2(z)$ is determined such that $\phi_1(z)$, $\phi_2(z)$ define a function $f(z)$, holomorphic in the region S' consisting of the points of C and the regions S_1', S_2'. Along C the functions $f(z)$, $\phi_1(z)$, $\phi_2(z)$ take equal values which are continuous with the values taken respectively in the regions S_1', S_2'. As may be seen, the substitution

$$z = x + iy = \Psi(\tau)$$

transforms the functions $\phi_1(z)$, $\phi_2(z)$ into the functions $F_1(\tau)$, $F_2(\tau)$ which are holomorphic in S_1, S_2, respectively, and along AB take equal values. Moreover, since S_2' is a reflection of S_1' with respect to C, S_2 is likewise a reflection of S_1 with respect to AB. The function $f(z)$ is likewise transformed into a function $F(\tau)$, holomorphic in the region S consisting of the points of AB and the regions S_1, S_2, such that it coincides with $F_1(\tau)$ in S_1 and with $F_2(\tau)$ in S_2 and along AB we have

$$F(t) = F_1(t) = F_2(t).$$

The function $F_2(\tau)$ is therefore an analytic continuation of $F_1(\tau)$ by Theorem II.

In a similar manner the substitution

$$\tau = \phi(z)$$

transforms the functions $F_1(\tau)$, $F_2(\tau)$, $F(\tau)$ into the functions $\phi_1(z)$, $\phi_2(z)$, $f(z)$, respectively, where, if $F_2(\tau)$ is an analytic continuation of $F_1(\tau)$, then $\phi_2(z)$ is likewise an analytic continuation of $\phi_1(z)$, such that $\phi_1(z)$, $\phi_2(z)$ take equal values with $f(z)$ for values of z along C.

Consequently, we see that whenever $\phi_2(z)$ is an analytic continuation of $\phi_1(z)$, then $F_2(\tau)$ is an analytic continuation of $F_1(\tau)$ and conversely. The necessary and sufficient condition that $F_2(\tau)$ is an analytic continuation of $F_1(\tau)$ leads then to the necessary and sufficient

condition that $\phi_2(z)$ is an analytic continuation of $\phi_1(z)$. Moreover, if $F_2(\tau)$ is obtained as an analytic continuation of $F_1(\tau)$ by means of Schwarz's method of reflection, then $\phi_2(z)$ is a continuation of $\phi_1(z)$ by reflection with respect to the arc C. But as we have seen the necessary and sufficient condition that $F_1(\tau)$ can be analytically continued by reflection upon AB is that $F_1(\tau)$ takes along AB real values which are continuous with the values taken by this function in S_1; that is, that $F_1(\tau)$ converges uniformly toward real values along AB. Accordingly the necessary and sufficient condition that $\phi_1(z)$ can be analytically continued across C by reflection is that $\phi_1(z)$ converges uniformly to real values along the arc C as the theorem requires.

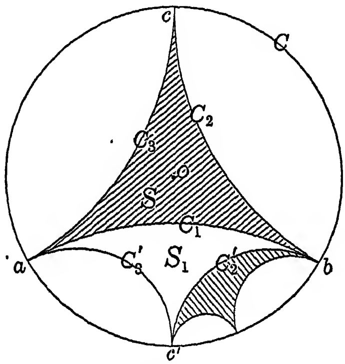

FIG. 94.

Let us now apply this method of analytic continuation by reflecting a given region with respect to an arc of a circle. Let C be any circle having its center at O, Fig. 94. Let the element $\phi_1(z)$ of the function $f(z)$ be defined for the region S bounded by three arcs C_1, C_2, C_3 of circles cutting the circle C at right angles and suppose that $\phi_1(z)$ converges uniformly to real values along C_1, C_2, C_3.

We shall now reflect the region S with respect to one of these arcs, say the arc C_1. In order to accomplish this we shall first show that the reflection of any point of S with respect to C_1 is the conjugate point obtained by geometric inversion of the given point with respect to the circle of which C_1 is an arc.

To show this consider the parametric equations of C_1. Suppose the center of C_1 to be the origin and its radius to be unity. Then any functions

$$x = \Psi_1(t), \qquad y = \Psi_2(t),$$

which satisfy the equation

$$x^2 + y^2 = 1$$

will answer our purpose. For example, we may put

$$x = \frac{1 - t^2}{1 + t^2}, \qquad y = \frac{2t}{1 + t^2}. \tag{13}$$

The functional relation (12) which maps the τ-plane upon the Z-plane is then

$$\begin{aligned} z = x + iy &= \frac{1 - \tau^2}{1 + \tau^2} + i\,\frac{2\tau}{1 + \tau^2} \\ &= \frac{1 + 2i\tau - \tau^2}{1 + \tau^2} = \frac{(1 + i\tau)^2}{1 + \tau^2} \\ &= \frac{1 + i\tau}{1 - i\tau}. \end{aligned} \tag{14}$$

Let z_1, z_2 be the points of the Z-plane corresponding respectively to the conjugate imaginary values $\tau_1 + i\tau_2$, $\tau_1 - i\tau_2$ of the τ-plane. We have then from (14)

$$z_1 = x_1 + iy_1 = \frac{1 + i(\tau_1 + i\tau_2)}{1 - i(\tau_1 + i\tau_2)} = \frac{1 + i\tau_1 - \tau_2}{1 - i\tau_1 + \tau_2},$$

and

$$z_2 = x_2 + iy_2 = \frac{1 + i(\tau_1 - i\tau_2)}{1 - i(\tau_1 - i\tau_2)} = \frac{1 + i\tau_1 + \tau_2}{1 - i\tau_1 - \tau_2};$$

whence, we obtain

$$x_2 + iy_2 = \frac{1}{x_1 - iy_1}, \tag{15}$$

or

$$x_2 + iy_2 = \frac{x_1 + iy_1}{x_1^2 + y_1^2}.$$

Equating the real parts and the imaginary parts, we have

$$x_2 = \frac{x_1}{x_1^2 + y_1^2}, \qquad y_2 = \frac{y_1}{x_1^2 + y_1^2}. \tag{16}$$

By comparing with the equations of transformation given in Art. 38, it will be seen that the reflection of z_1 with respect to the arc C_1 is merely the conjugate point of z_1 with respect to C_1.

The foregoing conclusion gives us a convenient method for determining the region S_1 which is the reflection of S with respect to C_1 and hence for determining an analytic continuation of $\phi_1(z)$. Since C cuts C_1 at right angles, the circle C inverts into itself. The points a, b remain unchanged, and the point c goes over into c'. As C_2 likewise cuts the circle C at right angles, it inverts into a circle C_2' perpendicular to C and passing through c' and b. In a similar manner the curve C_3 goes over into the curve C_3' passing through the points a, c' and cutting the circle C at right angles. The region S is then reflected into the region S_1. Associating with each point of S_1 the conjugate imaginary value of $\phi_1(z)$, where z is the corresponding point in S, we have by Theorem II $\phi_2(z)$, an analytic continuation of $\phi_1(z)$.

In a similar manner $\phi_3(z)$ is an analytic continuation of the given element $\phi_1(z)$ by reflection with respect to C_2 and $\phi_4(z)$, by reflection with respect to C_3. Continuing this process it is possible to enlarge the region S originally given, so as to include in the limit the entire region bounded by C.

50. Analytic function. By the aid of the results of the preceding article concerning analytic continuation, we can formulate more exactly the definition of an analytic function. If we know the values of a function and its derivatives at any point α, then, as we have already seen, an element $\phi_1(z)$ of that function is uniquely determined. By analytic continuation we can extend the region in which the function is thus defined by determining other elements of the function and their corresponding regions. This extended region forms a connected region S within which a function is defined by means of its elements. If we now suppose the region S to be extended as far as possible by means of analytic continuation, then the corre-

sponding aggregate of elements fully defines a function $f(z)$ in S such that $f(z)$ is equal to each of its elements $\phi(z)$ for those values of z for which $\phi(z)$ is defined. The function $f(z)$ so defined is called a **monogenic analytic function,** or more briefly an analytic function. As it is impossible to further extend this region S, it is called the **region of existence** of the analytic function $f(z)$. The element from which the other elements are obtained by the process of analytic continuation is called the **primitive element** of the function, and the remaining elements become analytic continuations of it.

The region of existence consists of a continuum of inner points, each of which is a regular point of the function $f(z)$. The region of existence may extend over the entire finite portion of the complex plane. On the other hand, it is possible in the process of analytic continuation to encounter a closed curve beyond which the function can not be analytically continued. In such a case the curve is called a **natural boundary.** For example, in the function discussed in connection with Fig. 94 of the last article, the curve C constitutes a natural boundary, since it is impossible to continue the function analytically across this curve. A portion of the complex plane into which the function can not be continued because of a natural boundary is called a **lacunary space.** Often, instead of a lacunary space, we encounter a set of points, not constituting a continuum, which can not be included in the region of existence. Such points are not regular points of the function, and hence they must be classed as singular points. The various classes of singularities of single-valued analytic functions will be more fully discussed in the following article.

Since power series may be used as a means of analytic continuation, it follows that an analytic function may also be defined as one that is developable, except in the neighborhood of singular points, by Taylor's expansion. It is to be noted also that a single-valued analytic function of a complex variable is uniquely determined throughout its region of existence as soon as its values in the neighborhood of any regular point of that region are given. The particular method of analytic continuation employed in extending the region from the neighborhood of the given point to the region of existence of the function thus determined is a matter of indifference. Moreover, any two analytic functions are equal for all values of z in this region of existence if they have a common element.

An important distinction between functions of a complex variable and those of a real variable may be noted. If a function of a com-

plex variable has a derivative at each point of a given region S, then at all points of S it has derivatives of every order, and in the neighborhood of any point of S the function is represented by the Taylor series to which it gives rise. On the other hand, it does not follow that if a function of a real variable has at each point of an interval derivatives of every order that the Taylor's expansion derived from the function represents that function for all values of the variable in the neighborhood of the point at which the derivatives are taken. The following example will serve as an illustration.

Ex. 1. Given the function $f(x) = e^{-\frac{1}{x^2}}$, $x \neq 0$, where $f(0) = 0$. Let it be required to determine the interval of equivalence of this function and the power series obtained from this function by expanding it in a Maclaurin series.

We have, for $x \neq 0$,

$$f(x) = e^{-\frac{1}{x^2}},$$

$$f'(x) = D_x e^{-\frac{1}{x^2}} = 2\frac{e^{-\frac{1}{x^2}}}{x^3},$$

$$f''(x) = D_x 2\frac{e^{-\frac{1}{x^2}}}{x^3} = (4 - 6x^2)\frac{e^{-\frac{1}{x^2}}}{x^6},$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$f^{(n)}(x) = G(x)\frac{e^{-\frac{1}{x^2}}}{x^{3n}},$$

where $G(x)$ is a polynomial in x.

For $x = 0$ we have by use of the limit *

$$\underset{x \doteq 0}{L} \frac{e^{-\frac{1}{x^2}}}{x^k} = \underset{y^2 = \infty}{L} \frac{(y^2)^{\frac{k}{2}}}{e^{y^2}} = 0,$$

$$f'(0) = \underset{\Delta x \doteq 0}{L} \frac{f(\Delta x) - f(0)}{\Delta x} = \underset{\Delta x \doteq 0}{L} \frac{e^{-\frac{1}{(\Delta x)^2}}}{\Delta x} = 0,$$

$$f''(0) = \underset{\Delta x \doteq 0}{L} \frac{f'(\Delta x) - f'(0)}{\Delta x} = \underset{\Delta x = 0}{L} 2\frac{e^{-\frac{1}{(\Delta x)^2}}}{(\Delta x)^4} = 0,$$

$$f'''(0) = \underset{\Delta x \doteq 0}{L} \frac{f''(\Delta x) - f''(0)}{\Delta x} = \underset{\Delta x \doteq 0}{L} \{4 - 6(\Delta x)^2\}\frac{e^{-\frac{1}{(\Delta x)^2}}}{(\Delta x)^7} = 0,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$f^{(n+1)}(0) = \underset{\Delta x \doteq 0}{L} G(\Delta x)\frac{e^{-\frac{1}{(\Delta x)^2}}}{(\Delta x)^{3n+1}} = 0.$$

* See Stolz, *Differential-und Integralrechnung*, p. 76, also p. 81, Ex. 3.

The expansion derived from the given function is a power series, each term of which is zero. This series defines, in the interval of convergence, a function $\phi(x) = 0$; that is, the function defined by the series is represented geometrically by the X-axis. On the other hand, the given function $y = f(x)$ is represented by the curve C tangent to the X-axis at the origin. This curve is symmetrical

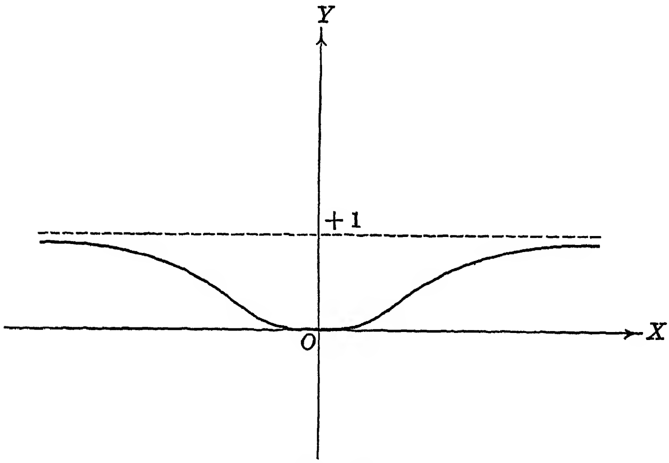

FIG. 95.

with respect to the Y-axis, and has the line $y = 1$ as an asymptote as shown in Fig. 95. It follows that the function that gave rise to the series is represented by that series in only one point, namely $x = 0$.

The reason for the distinction pointed out between functions of a complex variable and those of a real variable is, so far as the particular function discussed is concerned, that while $e^{-\frac{1}{z^2}}$ is an analytic function, the point $z = 0$ is not a regular point, since the derivative with respect to the complex variable z does not exist at the origin; although in the realm of real variables the corresponding derivative does exist. Consequently, the point $z = 0$ does not belong to the region of existence in the complex plane.

It is of importance to point out in this connection the distinction between an analytic function as defined and an **analytic expression.** The notion of an analytic function implies a definite correspondence between the z-points and the w-points of the complex plane. This relation may have different forms of expression in different parts of the plane. An analytic expression on the other hand is the result obtained by performing upon the independent variable the analytic

operations of addition, subtraction, multiplication, division, integration, etc., including the general process of taking the limit. It leads to a formal expression of the relation between z and w. This analytic expression, however, may define for different regions of the plane elements of different analytic functions. The following illustrations will make clear the distinction.

Consider the analytic expression

$$E(z) = \frac{z}{1-z^2} + \frac{z^2}{1-z^4} + \frac{z^4}{1-z^8} + \cdots .$$

This series converges * for all values of z except for values upon the unit circle about the origin. Within this circle the series converges to the limit

$$f_1(z) = \frac{z}{1-z}.$$

For values of z exterior to the unit circle the series converges to the limit

$$f_2(z) = \frac{1}{1-z}.$$

Hence, for $|z| < 1$, $E(z)$ may be considered as an element of the analytic function $\frac{z}{1-z}$, and for $|z| > 1$ it may be considered as an element of the analytic function $\frac{1}{1-z}$. In either case the element $E(z)$ may be analytically continued over the entire finite plane with the exception of the point $z = 1$, but in the one case the resulting analytic function is $\frac{z}{1-z}$, while in the other it is $\frac{1}{1-z}$.

As another illustration, suppose we have two detached regions S_1, S_2. Let $\phi_1(z)$, $\phi_2(z)$ be elements of two distinct analytic functions $f_1(z)$, $f_2(z)$. Suppose $\phi_1(z)$ to be defined for S_1, within which it is holomorphic, and along the boundary C_1 of S_1 let it converge uniformly. Let $\phi_2(z)$ be defined in a similar manner for S_2 and along its boundary C_2. Consider the analytic expression

$$E(z) = \frac{1}{2\pi i}\int_{C_1} \frac{\phi_1(t)\,dt}{t-z} + \frac{1}{2\pi i}\int_{C_2} \frac{\phi_2(t)\,dt}{t-z}. \tag{1}$$

It follows from Cauchy's integral formula that for values of z within S_1 we have

$$\phi_1(z) = \frac{1}{2\pi i}\int_{C_1} \frac{\phi_1(t)\,dt}{t-z}.$$

* See Bromwich, *Theory of Infinite Series*, p. 254, Ex. 4.

For values of z exterior to S_1, the integrand $\dfrac{\phi_1(t)}{t-z}$, considered as a function of t, is holomorphic in S_1 and hence by the Cauchy-Goursat theorem this integral vanishes. Similarly, the second integral in (1) defines the element $\phi_2(z)$ for values of z within S_2 and vanishes for all values of z exterior to that region. Hence, the expression $E(z)$ is equal to $\phi_1(z)$ for values of z within C_1 and to $\phi_2(z)$ for values of z within C_2. It follows then that $E(z)$ defines an element of the analytic function $f_1(z)$ or $f_2(z)$ according as z lies within C_1 or C_2.

51. Singular points and zero points. We have defined a singular point of a function (Art. 14) as a point that is not a regular point of the function, but in every deleted neighborhood of which there are regular points. As we have seen in the previous article the singular points of an analytic function are to be considered as boundary points of its region of existence, and they may even form a closed curve constituting a natural boundary of such a function. If α is a singular point of an analytic function $f(z)$, then either the function has no derivative at the point α itself, or there are points in every neighborhood of α at which the function has no derivative. In either case the higher derivatives of the function can not exist at α, and hence the function does not permit of an integral power series development in the neighborhood of α. If we undertake by means of power series to continue analytically an element of an analytic function along an ordinary curve passing through a singular point, the circles of convergence within which the successive elements are defined grow gradually smaller as their centers approach the singular point; for, as the function is always holomorphic within these circles none of them can ever inclose the singular point itself.

The singular points of a single-valued analytic function may be classified as poles, or non-essential singular points, and essential singular points. The point $z = \alpha$ is a **pole, or non-essential singular point,** of the analytic function $f(z)$ if there exists a positive integral value of k such that the product

$$(z-\alpha)^k f(z)$$

is holomorphic in the neighborhood of α and different from zero for $z = \alpha$. The integer k is called the **order of the pole.**

Thus the point $z = 2$ is a pole of order 2 of the function

$$f(z) = \frac{3z^2+1}{(z-2)^2}$$

for, multiplying $f(z)$ by $(z-2)^2$ we obtain the function $3z^2+1$, which has the point $z=2$ as a regular point and is different from zero for $z=2$. If k is equal to one, the pole is often referred to as a **simple pole.**

If no finite value of k exists such that the singularity of the single-valued analytic function $f(z)$ at a point α is removed by multiplying by the factor $(z-\alpha)^k$, then α is said to be an **essential singular point** of $f(z)$. If a singular point can be inclosed in a circle, however small, having that point as a center and containing no other singular point of the given function, then the point is said to be an **isolated singular point.** An isolated essential singular point is one that may be inclosed in a circle containing no other essential singular point. It may, however, have an infinite number of poles in its neighborhood, as we shall see later. A point may, therefore, be an isolated essential singular point without being an isolated singular point.

The following theorem due to Riemann is important in establishing the character of a function at a point in the deleted neighborhood of which it is limited in absolute value and holomorphic.*

THEOREM I. *Let $f(z)$ be holomorphic in a given region S except at the point $z=\alpha$, where the behavior of the function is not known. If for all values of $z \neq \alpha$ in S we have*

$$|f(z)| < M,$$

where M is some finite positive number, then $f(z)$ approaches a definite limit A as z approaches α and $z=\alpha$ is a regular point of $f(z)$ or may be made so by assigning to $f(z)$ at the point α the value $f(\alpha)=A$.

Denote by C any ordinary curve lying wholly within S and inclosing only points of S, including the point α. Let z be any point within C other than α. About α as a center draw a circle γ lying wholly within C and having an arbitrarily small radius ρ. The radius ρ can then be so chosen that z lies exterior to γ. We have then from Theorem I, Art. 20,

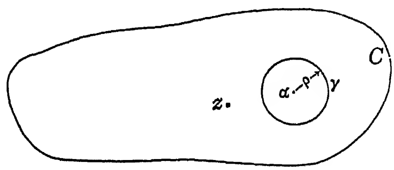

FIG. 96.

$$f(z)=\frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{t-z}+\frac{1}{2\pi i}\int_\gamma \frac{f(t)\,dt}{t-z}, \tag{1}$$

* See Osgood, *Bulletin of Amer. Math. Soc.*, June 1896, p. 298; also *Lehrbuch der Funktionentheorie*, Zweite Auflage, p. 310.

where the integral in each case is taken positively with reference to the region interior to C and exterior to γ. Since we have for values of t upon γ

$$|f(t)| < M, \qquad |t - z| \geqq |z - \alpha| - \rho,$$

$$\left|\frac{1}{2\pi i}\int_\gamma \frac{f(t)\,dt}{t-z}\right| < \frac{M}{2\pi i(|z-\alpha| - \rho)}\int_\gamma |dt| = \frac{M\rho}{|z-\alpha| - \rho};$$

for, $\int |dt| = 2\pi\rho$ is the length of the circle γ. As M and $|z - \alpha|$ are finite, the value of this integral is arbitrarily small for sufficiently small values of ρ. As this integral, however, does not vary with ρ, it must be equal to zero. Consequently, we have from (1)

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{t-z}, \tag{2}$$

which holds for any point z in S other than the point α.

By Theorem III, Art. 20, however, this integral defines a function $F(z)$, which is holomorphic everywhere within the region bounded by C, including the point $z = \alpha$ itself. The function $F(z)$ thus defined coincides with the given function $f(z)$ for all values of $z \neq \alpha$. Since $F(z)$ is holomorphic in S, we have

$$\underset{z \doteq \alpha}{L}\, F(z) = F(\alpha) = A.$$

If we now put

$$A = f(\alpha),$$

then $F(z)$ is identical with $f(z)$ for all values of z within C and consequently $z = \alpha$ is a regular point of the given function and

$$\underset{z \doteq \alpha}{L}\, f(z) = f(\alpha).$$

It follows as a result of this theorem that all isolated finite discontinuities of an analytic function may be removed by the proper definition of the function at the critical points. Consequently, we need not concern ourselves with the consideration of such discontinuities. This theorem also brings out an important distinction between functions of a complex variable and real functions of a real variable, which is best illustrated by an example.

Given the real function $f(x)$ of a real variable x defined by the relations

$$\begin{aligned} f(x) &= x \sin\frac{1}{x}, \text{ for } x \neq 0 \\ &= 0, \qquad \text{for } x = 0. \end{aligned}$$

This function is continuous at the origin, but has no derivative at that point although it possesses a derivative at every point in the neighborhood of the origin.*

A function $f(z)$ of a complex variable differs from a function of a real variable in that if $f(z)$ is holomorphic in the deleted neighborhood of any point α and continuous at α, then α is necessarily a regular point of the given function. In other words, a single-valued analytic function can not fail to have a derivative at an isolated point in the neighborhood of which it is continuous.

From Theorem II, Art. 49, it follows that if $f(z)$ is continuous in a given region S and if every point of S, with the exception of the points of an ordinary curve lying wholly within S, is a regular point of $f(z)$, then $f(z)$ is holomorphic in S. In other words, an analytic function of a complex variable can not have an isolated line of singular points in a region S in which it is continuous and, except for the points of this line, holomorphic. If an analytic function has all of the points of a curve as singular points, then that curve forms a portion of the boundary of the region of existence. If there exists a closed curve, every point of which is a singular point, then that curve constitutes a natural boundary of the function; for, in such a case the function can not be analytically continued beyond the curve.

We may state the following theorem concerning an analytic function.

THEOREM II. *Given an analytic function $f(z)$ which is not identically zero and which is holomorphic in the deleted neighborhood of $z = z_0$, then if $\underset{z=z_0}{L} f(z) = 0$ we can write $f(z)$ in the form*

$$(z - z_0)^k \phi(z),$$

where k is a positive integer and $\phi(z)$ is different from zero for $z = z_0$ and has this point as a regular point.

It follows from Theorem I that $f(z)$ is holomorphic in the neighborhood of z_0 and for z_0 we have $f(z_0) = 0$. We may then expand $f(z)$ in powers of $(z - z_0)$ by means of a Taylor series. This expansion is of the form

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \frac{f''(z_0)}{2!}(z - z_0)^2 + \cdots,$$

where, as we have seen, $f(z_0) = 0$. Not all of the derivatives can vanish; for, in that case the given function $f(z)$ would be identically zero for

* See Pierpont, *Theory of Functions of a Real Variable*, Vol. I, p. 225.

all points in the neighborhood of z_0 and hence throughout its region of existence. The first non-vanishing term must therefore contain the factor $(z - z_0)$ to some power, say the k^{th} power. We have then as the form of the expansion

$$f(z) = \alpha_k(z - z_0)^k + \alpha_{k+1}(z - z_0)^{k+1} + \cdots.$$

We may remove the factor $(z - z_0)^k$ from each term of the series and have

$$f(z) = (z - z_0)^k[\alpha_k + \alpha_{k+1}(z - z_0) + \cdots].$$

Since the series in the brackets is a power series it represents some function $\phi(z)$ which is holomorphic in the neighborhood of z_0. Moreover, $\phi(z_0) = \alpha_k$ is different from zero. Hence we have

$$f(z) = (z - z_0)^k\phi(z),$$

where $\phi(z)$ satisfies the conditions set forth in the theorem.

The point $z = z_0$ is said to be a **zero point of order** k of the analytic function $f(z)$, if there exists a positive real integral value of k such that the product

$$\frac{1}{(z - z_0)^k} \cdot f(z)$$

is holomorphic in the neighborhood of z_0 and different from zero for $z = z_0$. If $z = z_0$ is a regular point of the analytic function $f(z)$ and if $f(z_0) = 0$, then by the foregoing theorem $z = z_0$ is a zero point. By multiplying $f(z)$ by the factor $\frac{1}{(z - z_0)^k}$, where k is the order of the zero point, the vanishing point is removed.

That a theorem does not exist for real variables analogous to Theorem II for analytic functions of a complex variable is illustrated by the function

$$f(x) = e^{-\frac{1}{x^2}}, \qquad x \neq 0.$$

This function satisfies the conditions of Theorem II, stated with reference to the real domain in the deleted neighborhood of the origin; that is, it has all derivatives with respect to x in this deleted neighborhood, and moreover,

$$\underset{x \doteq 0}{L} f(x) = 0.$$

But we have the limit *

$$\underset{x \doteq 0}{L} \frac{f(x)}{x^k} = 0$$

* See Stolz, *Differential-und Integralrechnung*, Part I, p. 81.

for all values of k, and hence the zero point can not be removed by introducing the factor $\frac{1}{x^k}$, no matter how large k may be taken.

Between the zero points and the poles of an analytic function, there exists the following relation.

THEOREM III. *If $z = z_0$ is a pole of order k of the analytic function $f(z)$, then $\frac{1}{f(z)}$ is holomorphic in the neighborhood of z_0 and has a zero point of order k at z_0, and conversely.*

Since $z = z_0$ is a pole of order k of $f(z)$, we have from the definition of a pole

$$(z - z_0)^k f(z) = \phi(z), \tag{3}$$

where $\phi(z)$ is holomorphic in the neighborhood of z_0 and $\phi(z_0) \neq 0$. Hence in the neighborhood of z_0, $\frac{1}{\phi(z)} \equiv \Psi(z)$ is also holomorphic and can be expanded in a power series, the first term of which is a constant different from zero. From (3) we have

$$\frac{1}{f(z)} = (z - z_0)^k \cdot \frac{1}{\phi(z)} = (z - z_0)^k \cdot \Psi(z). \tag{4}$$

Since both $(z - z_0)^k$ and $\Psi(z)$ are holomorphic in the neighborhood of z_0 and $\Psi(z_0) \neq 0$, the last member of (4) is holomorphic in the neighborhood of z_0 and can be expanded in powers of $(z - z_0)$ beginning with the k^{th}. Hence $\frac{1}{f(z)}$ has a zero point of the order k, as stated in the theorem.

The converse of the foregoing theorem follows similarly; for, if $\frac{1}{f(z)}$ is holomorphic with a zero point of order k at z_0, we may write

$$\frac{1}{f(z)} = (z - z_0)^k F(z), \tag{5}$$

where $F(z)$ is holomorphic and different from zero for $z = z_0$. Consequently, we have

$$(z - z_0)^k f(z) = \frac{1}{F(z)},$$

where $\frac{1}{F(z)}$ is also holomorphic and different from zero for $z = z_0$. Hence z_0 is a pole of order k of the given function $f(z)$.

In the foregoing demonstration we have made use of the fact

that if a function is holomorphic and different from zero in the neighborhood of z_0, then the reciprocal function $\frac{1}{f(z)}$ is also holomorphic and different from zero in the neighborhood of z_0. If, on the other hand, the point z_0 is not a regular point of the function $f(z)$, then this point must be a zero point or an essential singular point of the reciprocal function according as it is a pole or an essential singular point of $f(z)$. For example, $z = 0$ is an essential singular point of $e^{-\frac{1}{z^2}}$ and is also an essential singular point of $e^{\frac{1}{z^2}}$.

By aid of Theorem III we can now establish the following theorem.

THEOREM IV. *If an analytic function $f(z)$ is holomorphic and different from zero in the deleted neighborhood of $z = z_0$, and if*

$$\underset{z \doteq z_0}{L} f(z) = \infty,$$

then the point $z = z_0$ is a pole of $f(z)$.

Since

$$\underset{z \doteq z_0}{L} f(z) = \infty,$$

we have at once

$$\underset{z \doteq z_0}{L} \frac{1}{f(z)} = 0.$$

As $f(z)$ is holomorphic in the deleted neighborhood of z_0, it follows that $\frac{1}{f(z)}$ is holomorphic and different from zero in the same deleted neighborhood. Hence, by Theorem II there exists a positive integer k such that

$$\frac{1}{f(z)} = (z - z_0)^k \phi(z),$$

where $\phi(z)$ is holomorphic in the neighborhood of z_0 and $\phi(z_0) \neq 0$. The function $\frac{1}{f(z)}$ has then a zero point at z_0, and consequently by Theorem III, $f(z)$ must have a pole at the same point. Hence the theorem.

As in the case of Theorems I, II, the analogous theorem for the realm of real variables does not exist, as the following illustration shows.

EX. 1. Show that Theorem IV does not hold for the following function (Fig. 97) of a real variable, namely:

$$f(x) = e^{\frac{1}{x^2}}, \qquad x \neq 0.$$

The conditions of Theorem IV are satisfied for real values of the variable in the deleted neighborhood of the origin. But as x approaches zero the product $x^k f(x)$ becomes infinite * for all values of k. Hence, the infinity of the function can not be removed by introducing the factor x^k no matter how large k be chosen.

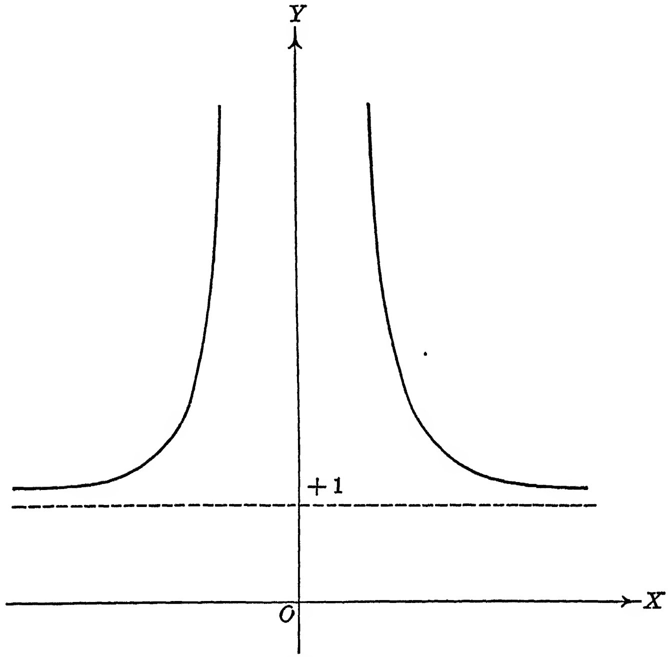

FIG. 97.

THEOREM V. *The zero points of an analytic function are isolated.*

Let z_0 be any zero point of the analytic function $f(z)$. We may then write

$$f(z) = (z - z_0)^k \phi(z), \tag{6}$$

where $\phi(z)$ is holomorphic in the neighborhood of z_0 and different from zero for $z = z_0$. Since $\phi(z)$ is continuous, we can then draw a circle C about z_0 as a center, within which $\phi(z)$ does not vanish. For any value of $z \neq z_0$ within C, $(z - z_0)^k$ is likewise different from zero. Consequently, within C there is no point other than z_0 at which $f(z)$ vanishes. The zero point z_0 is therefore isolated. But z_0 was any zero point of $f(z)$ and hence all such points are isolated.

* See Stolz, *Differential-und Integralrechnung*, p. 81.

The corresponding theorem for the poles of an analytic function may be stated as follows:

THEOREM VI. *The poles of an analytic function are isolated singular points.*

If an analytic function $f(z)$ has a pole at any point z_0, then by Theorem III $\frac{1}{f(z)}$ is holomorphic in the neighborhood of z_0 and has the value zero at z_0. But we have just seen (Theorem V) that the zero points of an analytic function are isolated. Consequently, the poles must also be isolated.

If the poles of an analytic function $f(z)$ have a limiting point, then the behavior of $f(z)$ in the neighborhood of that point is given by the following theorem.

THEOREM VII. *If $z = z_0$ is a limiting point of the poles of an analytic function $f(z)$, then $f(z)$ has an essential singularity at z_0.*

In every neighborhood of z_0 there are poles of the given function. The point $z = z_0$ can not then be a regular point of the function, and hence must be either a pole or an essential singular point. It can not be a pole, because as we have seen (Theorem VI) every pole is an isolated singular point. It must then be an essential singular point as the theorem states.

If an analytic function has an infinite number of poles, they must have at least one limiting point either in the finite region or at infinity. We now see that at this limiting point the function has an essential singularity. It follows then that an analytic function having no essential singularities can have but a finite number of poles.

We have seen that if $f(z)$ is holomorphic in the deleted neighborhood of z_0 and

$$\underset{z \doteq z_0}{L} f(z) = \infty,$$

then z_0 is a pole of $f(z)$. We shall now show that conversely an analytic function always becomes infinite as the variable approaches a pole; that is, we shall demonstrate the following theorem.

THEOREM VIII. *If the analytic function $f(z)$ has a pole at $z = z_0$, then the function $f(z)$ always becomes infinite as z approaches z_0 by any path; that is,*

$$\underset{z \doteq z_0}{L} f(z) = \infty.$$

Suppose that $f(z)$ has a pole of order k at z_0, then by the definition of a pole, we have

$$(z - z_0)^k f(z) = \phi(z),$$

where $\phi(z)$ is holomorphic in the neighborhood of z_0 and for $z = z_0$ is different from zero. For values of $z \neq z_0$, we have then

$$f(z) = \frac{\phi(z)}{(z - z_0)^k}.$$

As $\phi(z)$ is finite and continuous for $z = z_0$, then as z approaches z_0 by any path whatsoever $\frac{\phi(z)}{(z - z_0)^k}$ increases in absolute value without limit. Consequently, we may write

$$\underset{z \doteq z_0}{L} f(z) = \infty,$$

as the theorem requires.

Not only may an essential singular point of an analytic function appear as a limiting point of poles, but it may also be the limiting point of other essential singular points or it may appear as an isolated singular point of the function. For isolated essential singular points, that is essential singular points that are not the limiting points of other essential singular points, we have the following theorem.*

THEOREM IX. *If z_0 is an isolated essential singular point of $f(z)$, and β is any arbitrary number, real or complex, then z may be made to approach z_0 in such a manner that the corresponding values of $f(z)$ have the limiting value β.*

By hypothesis the point z_0 can not be the limiting point of other essential singular points of the function, although it may be the limiting point of poles of the function. Moreover, there can not exist a neighborhood of z_0, however small, such that at every point of it we have $f(z) = \beta$; for, in this case $f(z)$ would have by Theorem I at most a removable discontinuity at z_0, and hence this point could not be an essential singular point. Consider the function

$$F(z) = \frac{1}{f(z) - \beta}$$

for values of z in the neighborhood of z_0.

* This theorem, commonly attributed to Weierstrass, was doubtless first demonstrated by the Italian mathematician, Casorati.

See *Rend. Ist. Lomb.*, (2) I, 1868; also Vivanti-Gutzmer, *Theorie der eindeutigen analytischen Funktionen*, p. 130.

Either there exists a finite number M such that for all values of z in every neighborhood of z_0 we have

$$|F(z)| < M,$$

or there exists no such number M. We shall show that the first of these alternatives is impossible under the conditions of the theorem. In fact, if a finite number M can be found, then by Theorem I, z_0 is a regular point of $F(z)$. If in addition $F(z)$ is equal to zero for $z = z_0$, then z_0 is by Theorem III a pole of $f(z) - \beta$ and hence of $f(z)$. If, on the other hand, $F(z)$ is not equal to zero, then z_0 is a regular point of $f(z) - \beta$ and hence of $f(z)$. But either of these conclusions is a contradiction to the given hypothesis, for z_0 is by the conditions of the theorem an essential singular point of $f(z)$. It follows that there can exist no finite value of M such that for all values of z in every neighborhood of z_0 we have $|F(z)| < M$.

In every neighborhood of z_0 however small there are then points at which $|F(z)| > M$, however large M may be taken; that is, in every neighborhood of z_0 there are values of z for which $|F(z)| > \frac{1}{\epsilon}$, where ϵ is arbitrarily small. For all such values we have

$$|f(z) - \beta| < \epsilon.$$

We may then select a set of points

$$z_1, z_2, \ldots, z_n, \ldots,$$

having z_0 as a limiting point, such that

$$\underset{z_n \doteq z_0}{L} f(z_n) = \beta.$$

The foregoing theorem must not be understood to mean that in the neighborhood of an essential singular point $f(z)$ actually takes every value. Picard has shown,* however, that in the neighborhood of an essential singular point a single-valued analytic function takes all complex values with the exception of at most two and indeed an infinite number of times.

Thus far we have considered only those singularities of a single-valued function that occur at finite points of the complex plane. To determine the nature of the function in the neighborhood of the

* See *Mémoire sur les fonctions entières*, Ann. de l'École Normale, 1880; also *Traité d'analyse*, Vol. II, p. 121.

infinite point, we subject the variable z to the reciprocal transformation

$$z = \frac{1}{z'}$$

and examine the transformed function $\phi(z')$ for values of z' in the neighborhood of the point $z' = 0$. The given function $f(z)$ is said to have **a pole or an essential singularity at infinity** according as $z' = 0$ is a pole or an essential singular point of $\phi(z')$. The function (z) is said to have a **regular point at infinity** if $z' = 0$ is a regular point of the function $\phi(z')$.

In case the point $z = \infty$ is a regular point of the given function $f(z)$, then the transformed function $\phi(z')$ is holomorphic in the neighborhood of $z' = 0$. We can then expand $\phi(z')$ in a Maclaurin series and have

$$\phi(z') = \alpha_0 + \alpha_1 z' + \cdots + \alpha_n z'^n + \cdots .$$

Consequently, the expansion of $f(z)$ in the neighborhood of $z = \infty$, when this point is a regular point of the function, is of the form

$$f(z) = \alpha_0 + \frac{\alpha_1}{z} + \cdots + \frac{\alpha_n}{z^n} + \cdots .$$

It has been shown that if $f(z)$ is holomorphic in a given region S, then the integral $\int f(z)\, dz$ taken around any closed curve C lying wholly within S and inclosing only points of S must vanish. It is of interest in this connection to point out that this conclusion does not hold when the given region includes the point at infinity. For this case, we have the following theorem.

THEOREM X. *If C is an ordinary curve inclosing the point at infinity and lying within a given region S which likewise contains the point at infinity, then the integral $\int_C f(z)\, dz$ vanishes if $z^2 f(z)$ is holomorphic in S.*

Putting $z = \dfrac{1}{z'}$, we have

$$\int_C f(z)\, dz = -\int_\gamma z'^{-2}\, \phi(z')\, dz', \qquad \text{(Art. 22)}$$

where γ is the curve about the origin into which the curve C is mapped by the transformation $z = \dfrac{1}{z'}$. The given integral vanishes whenever

the integral $-\int_\gamma z'^{-2}\,\phi(z')\,dz'$ vanishes, that is, if $z'^{-2}\,\phi(z')$ is holomorphic in a region S' about the origin within which the curve γ lies. However, if $z'^{-2}\,\phi(z')$ is holomorphic in S', then $z^2 f(z)$ must be holomorphic in the corresponding region S about the point infinity and conversely. Hence the theorem.

THEOREM XI. *The circle of convergence of a power series passes through at least one singular point of the analytic function determined by the series.*

In the discussion of Taylor's series (Art. 48) it was pointed out that the power series in $(z - z_0)$ resulting from the expansion of a given function which is holomorphic in a region S converges and represents that function for all values of z within any circle that can be drawn about z_0 as long as it lies within S and incloses only points of S. That is, the size of the circle C (Fig. 87) within which the series is known to converge is limited only by the region S in which the given function is holomorphic. As we now know, that region S is restricted only by the presence of singular points of the analytic function $f(z)$ of which the given power series defines an element. Consequently, if the circle C is the circle of convergence of the Taylor series, then it must pass through at least one singular point of $f(z)$; that is, within every larger concentric circle there must be at least one such point; otherwise, a larger circle than C might be selected in determining a region within which the Taylor series converges and the series would then converge for points outside the circle C. This circle would not then be the circle of convergence as assumed.

If a function is holomorphic in a given region, it can be expanded in a Taylor's series for values of z in the neighborhood of any point of that region. The expression obtained for the coefficients of such an expansion enables us to establish the following theorem, due to Liouville.

THEOREM XII. *A single-valued analytic function which has no singularity either in the finite portion of the plane or at infinity reduces to a constant.*

If a function $f(z)$ has no singularity either in the finite region or at infinity, it follows that it is everywhere less in absolute value than some definite number M; for, otherwise, there would exist a point z_0, finite or infinite, in every neighborhood of which $f(z)$ would exceed

in absolute value all finite bounds, that is, would become infinite. The function admits of a Maclaurin expansion about the origin, namely

$$\alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots + \alpha_n z^n + \cdots \tag{6}$$

which converges and represents the function for all finite values of z.

The coefficient α_n is

$$\frac{1}{n\,!} f^{(n)}(0) = \frac{1}{2\pi i}\int_C \frac{f(z)\,dz}{z^{n+1}},$$

where C is a circle of radius ρ about the origin as a center, the value of ρ being taken as large as we please. Since

$$|f(z)| < M,$$

we may write

$$|\alpha_n| < \frac{1}{2\pi}\int_C \frac{M\,dz}{\rho^{n+1}} = \frac{1}{2\pi}\cdot\frac{M}{\rho^{n+1}}\cdot 2\pi\rho = \frac{M}{\rho^n}.$$

Inasmuch as ρ can be taken as large as we choose, it follows that

$$|\alpha_n| < \epsilon, \qquad n > 0,$$

where ϵ is an arbitrarily small positive number. Consequently, since α_n is a constant, we must have

$$\alpha_n = 0, \qquad n = 1, 2, 3, \ldots .$$

It follows from equation (6) that

$$f(z) = \alpha_0$$

for all values of z in the finite portion of the plane. Since the point $z = \infty$ is a regular point of $f(z)$, we have

$$f(z) = \underset{z \doteq \infty}{L} f(z) = \alpha_0.$$

It follows from the foregoing theorem that every single-valued analytic function which is not a constant must have at least one singular point either in the finite portion of the plane or at infinity.

52. Laurent's expansion. We have seen that, in the neighborhood of a regular point of an analytic function, it can be represented by a power series, but this method of representation does not hold in the neighborhood of a singular point of the analytic function. We shall now show that in the neighborhood of an isolated singular point z_0 we can expand an analytic function in a series having also negative powers of $(z - z_0)$. Such a series is not properly a power series, since a power series was defined as a series involving only positive integral powers. We shall, however, often refer to the series involving negative powers as a **power series with negative exponents** or a power

series in $\dfrac{1}{z - z_0}$. When the term power series is used without a qualifying phrase, we shall as heretofore understand it to mean a series involving only the positive powers of the variable.

In the derivation of Taylor's expansion (Art. 48), it was found to be valid within a region bounded by a single circle, provided there are no singular points of the given analytic function within the circle. Suppose we now consider a region S bounded by two concentric circles C_1, C_2 (Fig. 98) such that within S, $f(z)$ has no singular points and converges uniformly to finite values along each circle. There are no restrictions as to singular points exterior to C_1 or interior to C_2. Denote the common center of C_1, C_2 by z_0. To apply this method later to the expansion of a function in the neighborhood of a singular point, it is convenient to take the radius of C_2 arbitrarily small.

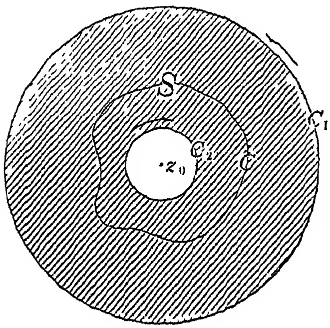

FIG. 98.

As in the consideration of Taylor's series, we shall base our discussion upon the fact that we can express the given analytic function $f(z)$ by means of the Cauchy integral formula. Since the given region S is bounded by two curves the integral must be taken over the entire boundary and hence along the two curves in the directions indicated in the figure. Taking, however, the integral along C_2 in a negative direction with respect to the region S, that is in a counter-clockwise direction, we have for any value of z in S

$$f(z) = \frac{1}{2\pi i}\int_{C_1}\frac{f(t)\,dt}{t - z} - \frac{1}{2\pi i}\int_{C_2}\frac{f(t)\,dt}{t - z}, \tag{1}$$

where t is taken along each of the curves C_1, C_2 in a counter-clockwise direction. Since z is any point in S, then for the first integral we have $|z - z_0| < |t - z_0|$. By Theorem III, Art. 20, this integral by itself defines a function $\phi(z)$ which is holomorphic for all values of z within C_1 and hence can be expanded in a power series in $(z - z_0)$ by means of Taylor's expansion. Such an expansion is of the form

$$\phi(z) = \alpha_0 + \alpha_1(z - z_0) + \alpha_2(z - z_0)^2 + \cdots + \alpha_n(z - z_0)^n + \cdots, \tag{2}$$

where we have

$$|z-z_0|<|t-z_0|, \quad \alpha_n=\frac{1}{2\pi i}\int_{C_1}\frac{f(t)\,dt}{(t-z_0)^{n+1}}, \qquad n=0, \quad 1,\ 2,\ \cdots.$$

The second integral also defines a function $\psi(z)$ which is holomorphic for values exterior to C_2, that is for $|z-z_0|>|t-z_0|$, the values of t being limited to values on C_2. To find a form of expansion for $\psi(z)$, we proceed in a manner similar to that used in the discussion of Taylor's series. We shall consider the function $\frac{1}{t-z}$, which occurs in the given integrand. We may write

$$\frac{1}{t-z}=\frac{1}{z-z_0}\left(\frac{z-z_0}{t-z}\right)=\frac{-1}{z-z_0}\left\{\frac{1}{1-\dfrac{t-z_0}{z-z_0}}\right\}$$

$$=-\frac{1}{z-z_0}-\frac{t-z_0}{(z-z_0)^2}-\frac{(t-z_0)^2}{(z-z_0)^3}-\cdots-\frac{(t-z_0)^{n-1}}{(z-z_0)^n}-\cdots. \tag{3}$$

This series, considered as a series in t, converges uniformly (Art. 45, Theorem I) for any constant value of z such that $|z-z_0|>|t-z_0|$, that is for any value of z exterior to C_2. The property of uniform convergence is not destroyed by multiplying each term of (3) by $f(t)$. We have then

$$\frac{f(t)}{t-z}=-\frac{f(t)}{z-z_0}-\frac{(t-z_0)f(t)}{(z-z_0)^2}-\frac{(t-z_0)^2f(t)}{(z-z_0)^3}-\cdots-\frac{(t-z_0)^{n-1}f(t)}{(z-z_0)^n}-\cdots.$$

Since this series converges uniformly, we may integrate it term by term, thus obtaining

$$\psi(z)=-\frac{1}{2\pi i}\int_{C_2}\frac{f(t)\,dt}{t-z}$$

$$=\frac{1}{2\pi i}\left\{\frac{1}{z-z_0}\int_{C_2}f(t)\,dt+\frac{1}{(z-z_0)^2}\int_{C_2}(t-z_0)f(t)\,dt+\cdots\right.$$

$$\left.+\frac{1}{(z-z_0)^n}\int_{C_2}(t-z_0)^{n-1}f(t)\,dt+\cdots\right\}.$$

The integrals in the second member of this equation determine the coefficients of the desired expansion of the second integral in (1). We may therefore write

$$\psi(z)=\alpha_{-1}(z-z_0)^{-1}+\alpha_{-2}(z-z_0)^{-2}+\cdots+\alpha_{-n}(z-z_0)^{-n}+\cdots, \tag{4}$$

where

$$|z-z_0|>|t-z_0|, \quad \alpha_{-n}=\frac{1}{2\pi i}\int_{C_2}(t-z_0)^{n-1}f(t)\,dt, \quad n=1,\ 2,\ \cdots.$$

Since the series (2) converges for all values of z within C_1 and the series (4) for all values of z exterior to C_2, it follows that both converge for values of z within S bounded by these two circles. Consequently for values of z within S, the given function $f(z)$ may be written as the sum of two functions $\phi(z)$, $\psi(z)$, the first of which can be expanded in a series involving the positive integral powers of $(z - z_0)$, and the second of which can be expanded in a series involving the negative integral powers of $(z - z_0)$; that is, we have

$$f(z) = \phi(z) + \psi(z) = \sum_{n=0}^{\infty} \alpha_n (z - z_0)^n + \sum_{n=1}^{\infty} \alpha_{-n} (z - z_0)^{-n}.$$

We may replace the two circles C_1, C_2 as paths of integration by a single path of integration. This path of integration may be any ordinary closed curve C lying within S, and inclosing C_2 since each of the circles C_1, C_2 may be deformed into C without passing over a singular point of the integrand. The coefficients of the two series (2) and (4) may then be expressed in terms of the integrals taken over the curve C. We have then the following theorem.

THEOREM I. *If $f(z)$ is holomorphic in the annular region S bounded by two concentric circles about a given point z_0, then within this region $f(z)$ can be represented by a series of the form*

$$\sum_{-\infty}^{\infty} \alpha_n (z - z_0)^n, \tag{5}$$

where

$$\alpha_n = \frac{1}{2\pi i} \int_C (t - z_0)^{-n-1} f(t)\, dt,$$

and C is any ordinary curve lying wholly within S and inclosing the inner circle.

The series (5) is known as **Laurent's series.** While there may be an infinite number of terms of the series corresponding to negative values of n, on the other hand only a finite number of such terms may appear in the expansion, the number depending as we shall see upon the character of the function $f(z)$ at the point z_0. By aid of the foregoing theorem we can now represent a single-valued analytic function in the deleted neighborhood of an isolated singular point by means of a series involving the positive and negative powers of the variable; for, if z_0 is such a singular point, then by making the radius of C_2 sufficiently small but different from zero we can include in the

region S any point in the deleted neighborhood of z_0. Hence, while $\sum_{n=0}^{\infty} \alpha_n(z - z_0)^n$ converges for all values of z within C_1, $\sum_{n=1}^{\infty} \alpha_{-n}(z - z_0)^{-n}$ converges for all values of z within C_1 except $z = z_0$.

As has already been pointed out, the nature of a singular point of an analytic function is fully determined by the behavior of the function in the deleted neighborhood of that point. The Laurent expansion of the function also determines the character of the singularity. For example, if $z = z_0$ is a pole of order k of the analytic function $f(z)$, then we are able to remove the singularity by multiplying $f(z)$ by the factor $(z - z_0)^k$. Hence there are k terms in the Laurent expansion having negative exponents; that is, the expansion is of the form

$$f(z) = \frac{\alpha_{-k}}{(z - z_0)^k} + \frac{\alpha_{-k+1}}{(z - z_0)^{k-1}} + \cdots + \frac{\alpha_{-1}}{(z - z_0)} + \alpha_0 + \alpha_1(z - z_0) + \alpha_2(z - z_0)^2 + \cdots + \alpha_n(z - z_0)^n + \cdots . \tag{6}$$

That part of the expansion which indicates the character of the singularity, namely

$$\sum_{r=-1}^{-k} \alpha_r(z - z_0)^r,$$

is called the **principal part** of the expansion. In case of a pole of order k, it consists of k terms.

The Laurent expansion of a given analytic function in the neighborhood of an isolated singular point may be accomplished by direct application of Theorem I, but if the singular point z_0 is a pole of order k, we may write

$$f(z) = \frac{\phi(z)}{(z - z_0)^k},$$

where

$$\phi(z) = \alpha_{-k} + \alpha_{-k+1}(z - z_0) + \cdots, \quad \alpha_{-k} \neq 0,$$

whence

$$f(z) = \frac{\alpha_{-k}}{(z - z_0)^k} + \frac{\alpha_{-k+1}}{(z - z_0)^{k-1}} + \cdots + \frac{\alpha_{-1}}{(z - z_0)} + \alpha_0 + \alpha_1 (z - z_0) + \cdots .$$

Ex. 1. Expand the function

$$f(z) = \frac{4z^2 - 4z + 1}{z^3 - z^4}$$

in a series for values of z in the neighborhood of the origin.

This function can be written in the form

$$f(z) = \frac{\phi(z)}{z^3},$$

where

$$\phi(z) = \frac{4z^2 - 4z + 1}{1 - z} = -4z + \frac{1}{1 - z} = 1 - 3z + z^2 + z^3 + \cdots .$$

Hence, we have

$$f(z) = \frac{1}{z^3} - \frac{3}{z^2} + \frac{1}{z} + 1 + z + z^2 + z^3 + \cdots .$$

The terms $\frac{1}{z^3} - \frac{3}{z^2} + \frac{1}{z}$ are the principal part of the expansion of $f(z)$ in the neighborhood of the origin.

If $f(z)$ has an isolated essential singular point at z_0 which is not the limiting point of poles of $f(z)$, then there is no finite value of k such that $(z - z_0)^k f(z)$ is holomorphic in the neighborhood of that point, and hence the Laurent expansion has an infinite number of terms involving negative powers of $(z - z_0)$. The expansion is then of the form

$$f(z) = \cdots + \frac{\alpha_{-k}}{(z - z_0)^k} + \frac{\alpha_{-k+1}}{(z - z_0)^{k-1}} + \cdots + \frac{\alpha_{-1}}{z - z_0} + \alpha_0 + \alpha_1(z - z_0) + \ldots + \alpha_n(z - z_0)^n + \ldots ; \quad (7)$$

that is, the principal part of the expansion consists of an infinite number of terms.

In case z_0 is a regular point of the function, the Laurent expansion has no terms with negative exponents and hence becomes identical with the Taylor expansion.

If the point $z = \infty$ is a pole of order k of a given function $f(z)$, then the function $\phi(z')$ obtained by transforming $f(z)$ by the relation $z = \frac{1}{z'}$ must have a pole of order k at the origin. Its expansion is therefore of the form

$$\phi(z') = \frac{\alpha_{-k}}{z'^k} + \frac{\alpha_{-k+1}}{z'^{k-1}} + \cdots + \frac{\alpha_{-1}}{z'} + \alpha_0 + \alpha_1 z' + \cdots + \alpha_n z'^n + \cdots .$$

In the neighborhood of $z = \infty$ the expansion of $f(z)$ is therefore of the form

$$f(z) = \alpha_{-k}z^k + \alpha_{-k+1}z^{k-1} + \cdots + \alpha_{-1}z + \alpha_0 + \frac{\alpha_1}{z} + \cdots + \frac{\alpha_n}{z^n} + \cdots ,$$

the first k terms constituting the principal part.

As with Taylor's expansion, the question arises as to whether an analytic function is uniquely represented by means of a Laurent series. In this connection we may well consider the following theorem.

THEOREM II. *If in an annular region S a given function $f(z)$ permits of an expansion of the form*

$$f(z) = \sum_{-\infty}^{\infty} \alpha_n (z - z_0)^n,$$

then the coefficients of this expansion are given by the relation

$$\alpha_n = \frac{1}{2\pi i}\int_C (t - z_0)^{-n-1} f(t)\, dt;$$

that is, there is but one such expansion possible.

As in the discussion of Theorem I, let the region S be bounded by the circles C_1, C_2, having the point z_0 as a center. We assume the existence of an expansion of $f(z)$ in a series as stated in the theorem. Denote by C any circle concentric with C_1, C_2 and lying between the two, and let the value of the variable along C be denoted by t. The given series converges along C and expressed in powers of $(t - z_0)$ is

$$f(t) = \cdots + \frac{\alpha_{-n}}{(t - z_0)^n} + \cdots + \frac{\alpha_{-2}}{(t - z_0)^2} + \frac{\alpha_{-1}}{(t - z_0)} + \alpha_0 + \alpha_1 (t - z_0) + \cdots + \alpha_n (t - z_0)^n + \cdots. \quad (8)$$

This series converges uniformly (Theorem I, Art. 45) and hence may be integrated term by term along C. Before doing so, however, let us multiply the terms of the series by the factor $\frac{1}{(t - z_0)^{n+1}}$. Remembering that

$$\int_C (t - z_0)^n\, dt = 0, \quad n \neq -1, \quad \text{(Exs. 1, 2, Art. 18)}$$
$$= 2\pi i, \quad n = -1, \quad (9)$$

it follows that the integrals of all of the terms of the series vanish except one, namely the term involving $\frac{1}{t - z_0}$. We have then as the result of the integration

$$\int_C \frac{f(t)\, dt}{(t - z_0)^{n+1}} = 2\pi i \alpha_n,$$

whence, we have

$$\alpha_n = \frac{1}{2\pi i}\int_C (t - z_0)^{-n-1} f(t)\, dt, \quad (10)$$

which establishes the theorem.

It does not follow from what has been said that $f(z)$ may not have different Laurent expansions in different circular regions. For example, suppose we have two regions (Fig. 99) one bounded by the circles C_1C_2 and the other by C_2C_3, where C_2 has upon it a singular point of the given function. In each of these regions there is an expansion in a Laurent series, but the two expansions in such a case are not identical. This condition is illustrated by the following example.

Ex. 2. Given the function

$$f(z) = \frac{1}{z^2 - 3z + 2} = \frac{1}{z-2} - \frac{1}{z-1}. \tag{11}$$

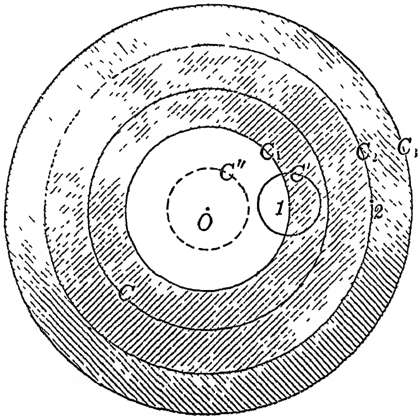

FIG. 99.

This function has two poles, namely, $z = 1$, $z = 2$. Within the circle about the origin passing through the point $z = 1$, the function can be represented by a Maclaurin series. The resulting series is

$$\frac{1}{2} + \frac{3}{4}z + \frac{7}{8}z^2 + \frac{15}{16}z^3 + \cdots + \frac{2^{n+1}-1}{2^{n+1}}z^n + \cdots,$$

which converges and represents the given function for $|z| < 1$.

If we take $1 < |z| < 2$, we must use Laurent's expansion. The coefficients of the series are given by

$$\alpha_n = \frac{1}{2\pi i}\int_C t^{-n-1}f(t)\,dt, \tag{12}$$

where C is any circle about the origin lying between C_1 and C_2.

Putting for $f(t)$ its value from (11), the integrand in (12) becomes

$$\frac{1}{t^{n+2} - 2t^{n+1}} - \frac{1}{t^{n+2} - t^{n+1}}.$$

Upon decomposing each of these fractions into partial fractions, we have for $n \geqq 0$

$$\frac{1}{2^{n+1}}\left\{\frac{1}{t-2} - \frac{1}{t} - \frac{2}{t^2} - \frac{4}{t^3} - \cdots - \frac{2^n}{t^{n+1}}\right\}$$
$$-\left\{\frac{1}{t-1} - \frac{1}{t} - \frac{1}{t^2} - \frac{1}{t^3} - \cdots - \frac{1}{t^{n+1}}\right\}.$$

Replacing the integrand in (12) by these partial fractions, since

$$\int_C t^n\,dt = 0, \qquad n \neq -1,$$
$$= 2\pi i, \quad n = -1,$$

we have

$$\alpha_n = \frac{1}{2^{n+2}\pi i}\int_C \frac{dt}{t-2} - \frac{1}{2^{n+1}} - \frac{1}{2\pi i}\int_C \frac{dt}{t-1} + 1. \tag{13}$$

But $\dfrac{1}{t-2}$ is holomorphic within C and hence the integral $\displaystyle\int_C \frac{dt}{t-2}$ vanishes. To evaluate the second integral in (13) we deform the path C of integration into a small circle C' about the point $z = 1$ and put

$$t - 1 = \rho e^{i\theta},$$

where ρ is a constant and θ varies from 0 to 2π. We have then

$$\int_C \frac{dt}{t-1} = \int_{C'} \frac{dt}{t-1} = i\int_0^{2\pi} d\theta = 2\pi i.$$

Consequently from (13), we have

$$\alpha_n = -\frac{1}{2^{n+1}} - 1 + 1 = -\frac{1}{2^{n+1}}.$$

The terms of the required Laurent expansion corresponding to values of $n \geqq 0$ are then

$$-\frac{1}{2} - \frac{1}{4}z - \frac{1}{8}z^2 - \cdots - \frac{1}{2^{n+1}}z^n - \cdots. \tag{14}$$

For negative values of n, say $n = -k$, we have

$$\begin{aligned}\alpha_{-k} &= \frac{1}{2\pi i}\int_C t^{k-1}f(t)\,dt \\ &= \frac{1}{2\pi i}\left[\int_C \frac{t^{k-1}\,dt}{t-2} - \int_C \frac{t^{k-1}\,dt}{t-1}\right].\end{aligned}$$

But the integrand in the first integral is holomorphic within C and hence the integral vanishes. We have by deforming C into C' and putting as before

$$t - 1 = \rho e^{i\theta},$$

$$\begin{aligned}\therefore \alpha_{-k} &= -\frac{1}{2\pi i}\int_0^{2\pi} \frac{(1+\rho e^{i\theta})^{k-1} i\rho e^{i\theta}\,d\theta}{\rho e^{i\theta}} \\ &= -\frac{i}{2\pi i}\int_0^{2\pi} (1+\rho e^{i\theta})^{k-1}\,d\theta \\ &= -\frac{i}{2\pi i}\int_0^{2\pi} d\theta = -1.\end{aligned}$$

Hence, each term in that portion of the Laurent expansion having negative exponents has the coefficient -1. The complete expansion is therefore

$$\cdots - \frac{1}{z^n} - \cdots - \frac{1}{z^2} - \frac{1}{z} - \frac{1}{2} - \frac{1}{4}z - \frac{1}{8}z^2 - \cdots - \frac{1}{2^{n+1}}z^n - \cdots. \tag{15}$$

It is evident that the given function has no finite singular point exterior to the circle about the origin and passing through the point $z = 2$. The expansion of the function about the point $z = \infty$ will then hold for this entire region. By putting

$$z = \frac{1}{z'},$$

the entire region exterior to the circle C_2 through $z = 2$ inverts into the region about the origin and lying within the circle C'' whose radius is $\frac{1}{2}$. The transformed function is

$$\phi(z') = \frac{z'}{1-2z'} - \frac{z'}{1-z'}.$$

Within the circle C'' the function $\phi(z')$ is holomorphic and may be expanded in a Maclaurin series, giving

$$\phi(z') = z'^2 + 3z'^3 + 7z'^4 + \cdots.$$

Replacing z' by $\frac{1}{z}$, we have as the expansion of the given function for values of z exterior to the circle C_2,

$$\frac{1}{z^2} + \frac{3}{z^3} + \frac{7}{z^4} + \cdots + \frac{2^{n-1}-1}{z^n} + \cdots.$$

The same result would have been obtained had we expanded the function by computing the coefficients by aid of the formula given in Theorem II, where the path of integration is any circle C_3 about the origin and lying exterior to the concentric circle through $z = 2$.

53. Residues. We have seen that the integral $\int f(z)\,dz$ vanishes when taken around the boundary C of a region S, provided $f(z)$ is holomorphic in the open region S and at least converges uniformly to its values along C. Let us now consider the effect upon this integral when S contains one isolated singular point of $f(z)$. Before doing so, we introduce the following definition.

If the points of S, with the exception of at most the point z_0, are regular points of $f(z)$ and C is any closed curve about z_0 and lying wholly within S and containing only points of S, then the integral

$$\frac{1}{2\pi i}\int_C f(z)\,dz$$

taken in the positive direction is called the **residue** of $f(z)$ at z_0.

Suppose the point z_0 is a pole of the given function. We have then the following theorem.

THEOREM I. *If $f(z)$ is holomorphic in a given finite region S except at z_0, where it has a pole, then the residue of $f(z)$ at z_0 is equal to the coefficient of $(z - z_0)^{-1}$ in the expansion of $f(z)$ in powers of $(z - z_0)$.*

Let the pole at z_0 be of order k. Then the Laurent expansion of the function in powers of $(z - z_0)$ is of the form

$$\begin{aligned} f(z) &= \frac{\alpha_{-k}}{(z-z_0)^k} + \frac{\alpha_{-k+1}}{(z-z_0)^{k-1}} + \cdots + \frac{\alpha_{-1}}{z-z_0} + \alpha_0 + \alpha_1(z-z_0) + \cdots \\ &= \frac{\alpha_{-k}}{(z-z_0)^k} + \frac{\alpha_{-k+1}}{(z-z_0)^{k-1}} + \cdots + \frac{\alpha_{-1}}{z-z_0} + \phi(z), \end{aligned}$$

where $\phi(z)$ is holomorphic in the neighborhood of z_0, say within and

upon a circle C having z_0 as a center. Taking C as the path of integration, we have

$$\int_C f(z)\,dz = \alpha_{-k}\int_C (z-z_0)^{-k}\,dz + \cdots$$
$$+ \alpha_{-1}\int_C (z-z_0)^{-1}\,dz + \int_C \phi(z)\,dz. \tag{1}$$

The integral $\int_C \phi(z)\,dz$ vanishes, since $\phi(z)$ is holomorphic in the closed region bounded by C. To evaluate the remaining integrals we make use of the relations

$$\begin{aligned}\int_C (z-z_0)^n\,dz &= 0, \qquad n \neq -1,\\ &= 2\pi i, \qquad n = -1.\end{aligned} \tag{2}$$

Consequently, we have from (1)

$$\int_C f(z)\,dz = 2\pi i\,\alpha_{-1},$$

whence

$$\alpha_{-1} = \frac{1}{2\pi i}\int_C f(z)\,dz, \tag{3}$$

which establishes the theorem, since by definition the second member of this equation is the residue.

The value of the residue of an analytic function at a pole is zero if the coefficient α_{-1} is zero in the expansion of the function. For example, the function

$$f(z) = \frac{1}{z^3},$$

has a pole of order three at the origin, yet the residue,

$$\frac{1}{2\pi i}\int_C \frac{dz}{z^3}$$

is zero.

The foregoing theorem gives the residue when there is a single isolated pole in the given region S. If there are a finite number of poles in S, we have the following theorem.

THEOREM II. *Given a function $f(z)$ which is holomorphic in a region S with the exception of a finite number of poles, and let C be any ordinary curve lying wholly within S and inclosing all of the given poles. Then $\int_C f(z)\,dz$ taken in a positive direction is equal to $2\pi i$ times the sum of the residues of $f(z)$ at these poles.*

Suppose the poles of $f(z)$ to be $z_1, z_2, \ldots, z_k, \ldots, z_n$. About each of these points as a center draw an arbitrarily small circle lying wholly within the region bounded by C. Denote these circles by $C_1, C_2, \ldots, C_k, \ldots, C_n$. Then by Theorem VI, Art. 19, we have

$$\int_C f(z)\,dz = \sum_{k=1}^{n} \int_{C_k} f(z)\,dz. \tag{4}$$

But as we have seen $\int_{C_k} f(z)\,dz$ is equal to $2\pi i$ times the residue of $f(z)$ at z_k and is given by the coefficient of $(z - z_k)^{-1}$ in the Laurent expansion of $f(z)$ in powers of $(z - z_k)$. The relation given in (4) therefore establishes the theorem.

A closed curve C may be regarded as the boundary of either of two regions, one finite and the other inclosing the point at infinity. It is readily seen that the relation (4) holds when C is regarded as bounding the outer region, as well as in the case just considered. Theorem II is still valid then when $\int_C f(z)\,dz$ is taken over a curve inclosing the point $z = \infty$. However, when the point $z = \infty$ is a pole the residue at that point is not given by (3). For this case we have the following theorem.

THEOREM III. *If the analytic function $f(z)$ has a pole at $z = \infty$, then the residue of $f(z)$ at that point is the negative of the coefficient of z^{-1} in the expansion of $f(z)$ for values of z in the neighborhood of $z = \infty$.*

Putting $z = \dfrac{1}{z'}$, we denote the transformed function by $\phi(z')$. As $z = \infty$ is a pole, say of order k, of the given function $f(z)$ then $z' = 0$ is a pole of the same order of $\phi(z')$. Expanding $\phi(z')$ in a Laurent series for values in the neighborhood of the origin we have

$$\phi(z') = \frac{\alpha_{-k}}{z'^k} + \frac{\alpha_{-k+1}}{z'^{k-1}} + \cdots + \frac{\alpha_{-1}}{z'} + \alpha_0 + \alpha_1 z' + \cdots + \alpha_n z'^n + \cdots. \tag{5}$$

Replacing z' by $\dfrac{1}{z}$, we have the expansion of $f(z)$ in the neighborhood of $z = \infty$, namely:

$$f(z) = \alpha_{-k}z^k + \alpha_{-k+1}z^{k-1} + \cdots + \alpha_{-1}z + \alpha_0 + \frac{\alpha_1}{z} + \cdots + \frac{\alpha_n}{z^n} + \cdots.$$

$$= \alpha_{-k}z^k + \alpha_{-k+1}z^{k-1} + \cdots + \alpha_{-1}z + \alpha_0 + \frac{\alpha_1}{z} + F(z), \tag{6}$$

where $z^2F(z)$ is holomorphic in the neighborhood of $z = \infty$. In determining the residue of $f(z)$ at the point $z = \infty$ the integral defining a residue is to be taken around an arbitrarily large circle C about the origin in a clockwise direction. The resulting integral is then the negative of the integral taken around C in a positive or counterclockwise direction. The integral $\int_C F(z)\,dz$ vanishes by Theorem X, Art. 51. We have then from (6) by aid of the relations given in (2), it being understood that the integral is taken in a clockwise direction around C,

$$-\int_C f(z)\,dz = 2\pi i\alpha_1,$$

or

$$\frac{1}{2\pi i}\int_C f(z)\,dz = -\alpha_1.$$

Consequently, from the definition of a residue, it follows that the residue at ∞ of an analytic function having a pole at infinity is the negative of the coefficient of z^{-1} in the expansion of the function in the neighborhood of the point $z = \infty$, as stated in the theorem.

It is of interest to observe that a function can have the point $z = \infty$ as a regular point and still have a residue at that point different from zero. For example, the function

$$f(z) = \alpha_0 + \frac{\alpha_1}{z}$$

is holomorphic in the neighborhood of $z = \infty$, yet it has a residue $-\alpha_1$ at that point.

The theory of residues is of value in the discussion of some of the important properties of analytic functions, as we shall see in the succeeding articles. It may also be applied with advantage to the evaluation of certain integrals of functions of a real variable, as we shall now show. First of all, we shall show how we may employ the results of our discussion of residues to evaluate an integral of the form

$$\int_{-\infty}^{\infty} f(x)\,dx,$$

where $f(x)$ is the quotient of two polynomials. In order that this integral shall have a significance, we make the assumption that the denominator is of degree at least two higher than the numerator.* For the sake of simplicity, we shall also assume that the denominator has no real roots.

* See Pierpont, *Theory of Functions of Real Variables*, Vol. I, Art. 635.

Let us consider then a function $f(z)$ which is holomorphic along the axis of reals, and with the exception of at most a finite number n of poles, holomorphic in the finite upper half of the complex plane. Consider now a region S bounded by a curve consisting of a segment of the axis of reals and a semicircle C about the origin, lying in the upper half plane and having the radius ρ. We select the value of ρ so that the poles of $f(z)$ already mentioned shall all lie in S. Denoting the residue of the given function at each pole z_k by R_k, then by Theorem II, we have upon integrating about the contour of S

$$\int_{-\rho}^{\rho} f(z)\,dz + \int_C f(z)\,dz = 2\pi i \sum R_k, \tag{7}$$

where $\int_{-\rho}^{\rho}$ denotes the integral along the axis of reals between $-\rho$ and ρ.

We shall first consider the limit

$$\underset{\rho=\infty}{L} \int_C f(z)\,dz.$$

Put $z = \rho e^{i\theta}$ and hence $\dfrac{dz}{z} = i\,d\theta$. We have, therefore,

$$\int_C f(z)\,dz = i\int_0^{\pi} zf(z)\,d\theta.$$

But, from 5, Art. 17, we have

$$\left|\int_C f(z)\,dz\right| = \left|\int_0^{\pi} zf(z)\,d\theta\right| \leqq \int_0^{\pi} \left|zf(z)\right| d\theta. \tag{8}$$

Let M denote the upper limit of $|zf(z)|$ upon the semicircle C. From (8), we have then

$$\left|\int_C f(z)\,dz\right| < M\int_0^{\pi} d\theta = \pi M.$$

If as ρ increases without limit, M approaches zero, we have

$$\underset{\rho=\infty}{L} \int_C f(z)\,dz = 0. \tag{9}$$

We now regard $f(z)$ as the quotient of two polynomials, where the denominator has no real roots and is of degree at least two higher than that of the numerator. Then the required conditions are satisfied, for $f(z)$ has a zero of at least order two at infinity and

$$\underset{\rho=\infty}{L} M = 0;$$

that is, the result expressed by (9) follows.

Moreover, under the conditions set forth in the statement of the problem each of the limits

$$\underset{\rho=\infty}{L} \int_0^{\rho} f(x)\,dx, \qquad \underset{\rho=\infty}{L} \int_0^{-\rho} f(x)\,dx$$

exists, and by passing to the limit we obtain from (7) the value of the integral $\int_{-\infty}^{\infty} f(z)\,dz$, namely

$$\int_{-\infty}^{\infty} f(z)\,dz = 2\pi i \sum_{k=1}^{n} R_k.$$

Ex. 1. Evaluate the integral $\int_{-\infty}^{\infty} \frac{dx}{(x^2+1)^3}$.

The function $f(z) = \frac{1}{(z^2+1)^3}$ has a pole at $z = i$. Expanding $f(z)$ in powers of $(z - i)$ we have

$$f(z) = \frac{i}{8(z-i)^3} - \frac{3}{16(z-i)^2} - \frac{3i}{16(z-i)} + \cdots.$$

Hence, the residue of $f(z)$ at $z = i$ is $-\frac{3i}{16}$.

We have then

$$\int_{-\infty}^{\infty} \frac{dx}{(x^2+1)^3} = -\frac{3i}{16} \cdot 2\pi i = \frac{3\pi}{8}.$$

The method of residues may be applied also in evaluating certain integrals of trigonometric functions when the integrals are taken between the limits 0 and 2π, as the following example will illustrate.

Ex. 2. Evaluate the integral $\int_0^{2\pi} \frac{d\theta}{2+\cos\theta}$.

Putting $z = e^{i\theta}$, we have

$$d\theta = \frac{dz}{iz}, \qquad \cos\theta = \frac{e^{i\theta}+e^{-i\theta}}{2} = \frac{z+\frac{1}{z}}{2},$$

whence we obtain

$$\int_0^{2\pi} \frac{d\theta}{2+\cos\theta} = \int_C \frac{\frac{1}{iz}}{\frac{1}{2}\left(z+4+\frac{1}{z}\right)}\,dz = \frac{2}{i}\int_C \frac{dz}{z^2+4z+1}.$$

The integrand $f(z)$ has poles at the points

$$z = -2 \pm \sqrt{3}.$$

The path C is determined by the substitution $z = e^{i\theta}$. As θ varies from 0 to 2π, z describes the unit circle about the origin. Of the two poles only one, namely, $z = -2 + \sqrt{3}$, falls within C. The residue of $f(z)$ at this pole is found to be $\frac{1}{2\sqrt{3}}$. Hence we have

$$\int_0^{2\pi} \frac{d\theta}{2+\cos\theta} = \frac{2}{i} \cdot 2\pi i \cdot \frac{1}{2\sqrt{3}} = \frac{2\pi}{3}\sqrt{3}.$$

54. Rational functions. Fundamental theorem of algebra. We have defined a rational integral function of z as a function of the form

$$G(z) = \alpha_0 + \alpha_1 z + \cdots + \alpha_n z^n,$$

and a rational fractional function as the quotient of two such functions having no common factor; that is,

$$f(z) = \frac{G_1(z)}{G_2(z)} = \frac{\alpha_0 + \alpha_1 z + \cdots + \alpha_n z^n}{\beta_0 + \beta_1 z + \cdots + \beta_m z^m}, \tag{1}$$

where if $n = m$, then the common value of n and m is said to be the degree of $f(z)$; otherwise the larger of the two numbers is the degree.

Every single-valued algebraic function is necessarily a rational function. For, if w is an algebraic function of z, then the two are connected by the irreducible equation

$$f_0(z)\, w^n + f_1(z)\, w^{n-1} + \cdots + f_n(z) = 0,$$

where $f_0(z), f_1(z), \ldots, f_n(z)$ are integral rational functions. If w is single-valued, this equation must be linear, and hence solving for w, we have

$$w = -\frac{f_1(z)}{f_0(z)};$$

that is, w is a rational function of z.

We shall now see how the two classes of rational functions are uniquely determined by the character of their singularities. As we have seen, every analytic function that is not a constant must have at least one singular point. For rational integral functions we may state the following theorem.

THEOREM I. *The necessary and sufficient conditions that a single-valued analytic function is a rational integral function are that it has no singular point in the finite portion of the complex plane and that it has a pole at infinity.*

We shall first show that the given conditions are necessary. Let $f(z)$ be a rational integral function; that is, let

$$f(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots + \alpha_n z^n,$$

where n is an integer. This function is holomorphic for all finite values of z, and hence has no singular points in the finite region of the complex plane. To determine the nature of the function for

$z = \infty$, we put $z = \frac{1}{z'}$, and examine the transformed function $\phi(z')$ for $z' = 0$. We have

$$\phi(z') = \alpha_0 + \frac{\alpha_1}{z'} + \frac{\alpha_2}{z'^2} + \cdots + \frac{\alpha_n}{z'^n},$$

which shows that $\phi(z')$ has a pole of order n at the origin, and hence $f(z)$ has a pole of the same order at the point $z = \infty$.

The given conditions are also sufficient. To show this we assume that $f(z)$ has no singular point of any kind in the finite region and that it has a pole, say of order n, at $z = \infty$. As we have seen in Art. 52 the expansion of $f(z)$ for values of z in the neighborhood of infinity is then of the form

$$f(z) = \alpha_n z^n + \alpha_{n-1} z^{n-1} + \cdots + \alpha_1 z + F(z),$$

where $F(z)$ has $z = \infty$ as a regular point. Since $f(z)$ is holomorphic everywhere in the finite portion of the plane, it follows that the same expansion holds for all finite values of z and that $F(z)$ must be holomorphic in the finite portion of the plane as well as in the neighborhood of infinity. The function $F(z)$ is therefore a constant (Theorem XII, Art. 51). Denoting this constant by α_0, we may write

$$f(z) = \alpha_n z^n + \alpha_{n-1} z^{n-1} + \cdots + \alpha_1 z + \alpha_0,$$

and hence $f(z)$ is a rational integral function as the theorem requires.

From this discussion it follows at once that a rational integral function is fully determined, except as to an additive constant, when the principal part of the expansion for the pole at infinity is known.

We can now establish the **fundamental theorem of algebra**, which may be stated as follows:

THEOREM II. *If $f(z)$ is a rational integral function, then the equation*

$$f(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots + \alpha_n z^n = 0 \tag{2}$$

has at least one root.

By Theorem I $f(z)$ has no singular points in the finite portion of the plane and has a pole at infinity. Putting $z = \frac{1}{z'}$, it follows that $\phi(z') = f\left(\frac{1}{z'}\right)$ has a pole at $z' = 0$. By Theorem III, Art. 51, $\frac{1}{\phi(z')}$ is then holomorphic in the neighborhood of the origin. Consequently, $\frac{1}{f(z)}$ must be holomorphic in the neighborhood of $z = \infty$.

But, as we have seen, every analytic function which is not a constant

must have at least one singular point either in the finite region or at infinity. Since $\frac{1}{f(z)}$ can not have a singular point at infinity, there must be at least one point in the finite region, say z_0, at which the function $\frac{1}{f(z)}$ has a singularity. This singular point can not be an essential singular point for in that case z_0 would not be a regular point of $f(z)$. Hence z_0 must be a pole of $\frac{1}{f(z)}$ and consequently a zero point of the given function $f(z)$. This establishes the theorem.

As a consequence of this theorem, it may be shown by the methods of elementary algebra that a rational integral function $f(z)$ may be written as the product of a constant times n linear factors, where n is the degree of $f(z)$. Consequently the equation (2) has n roots real or complex, and no more, each root being counted a number of times equal to its multiplicity.

We have the following condition that an analytic function is a rational fractional function.

THEOREM III. *The necessary and sufficient conditions that a single-valued analytic function is a rational fractional function are that it has at most a pole at infinity and that it has a finite number of poles but no essential singular points in the finite portion of the plane.*

The given conditions are necessary; for, if $f(z)$ is a rational fractional function, it can be written in the form given in (1), where $G_2(z)$ is not a constant. The finite singular points of $f(z)$ are the finite singular points of $G_1(z)$ and the finite singular points of $\frac{1}{G_2(z)}$, since there is by hypothesis no factor common to the two. By Theorem I $G_1(z)$ has no singular point in the finite portion of the plane. By Theorem II $G_2(z)$ has at least one zero point in the finite portion of the plane. Consequently, $\frac{1}{G_2(z)}$ has at least one pole in the finite portion of the plane. Since $G_2(z)$ can not have more than m zero points, $\frac{1}{G_2(z)}$ can not have more than m poles. Except in the neighborhood of the special points just mentioned $\frac{1}{G_2(z)}$ is holomorphic; for, in the neighborhood of every other finite point $G_2(z)$ is holomorphic and different from zero. It follows then that in the finite portion of the plane $f(z)$ has no essential singular points.

To determine the nature of the function $f(z)$ for $z = \infty$, we put $z = \dfrac{1}{z'}$ and have

$$f\left(\frac{1}{z'}\right) = \phi(z') = \frac{\alpha_0 + \alpha_1 \dfrac{1}{z'} + \cdots + \alpha_n \dfrac{1}{z'^n}}{\beta_0 + \beta_1 \dfrac{1}{z'} + \cdots + \beta_m \dfrac{1}{z'^m}}.$$

Consequently $\phi(z')$ has at $z' = 0$, and therefore $f(z)$ has at $z = \infty$, a pole, a regular point which is not a zero point, or a zero point according as $n > m$, $n = m$, $n < m$.

The given conditions are also sufficient. To show this let us assume that the poles of $f(z)$ in the finite region are at $\beta_1, \beta_2, \ldots, \beta_m$, and let us denote their orders by $k_1, k_2, \ldots, k_m$, respectively. Then from the definition of a pole, we have

$$G_1(z) = (z - \beta_1)^{k_1}(z - \beta_2)^{k_2} \ldots (z - \beta_m)^{k_m} f(z), \tag{3}$$

where $G_1(z)$ is holomorphic over the entire finite region and is different from zero for $z = \beta_1, \beta_2, \ldots, \beta_m$. However, if $G_1(z)$ is not a constant, it must have at least one singular point, and since this can not be a finite point, it must be the point at infinity. Moreover, this singular point must be a pole, for otherwise $f(z)$ would also have an essential singularity at infinity. It follows from Theorem I that $G_1(z)$ is either a constant or a rational integral function.

From (3) we have

$$f(z) = \frac{G_1(z)}{(z - \beta_1)^{k_1}(z - \beta_2)^{k_2} \ldots (z - \beta_m)^{k_m}} \equiv \frac{G_1(z)}{G_2(z)},$$

where $G_1(z)$ and $G_2(z)$ have no common factor. Consequently, $f(z)$ is a rational function as stated in the theorem.

A single-valued function $f(z)$ having a finite number of poles but no essential singular points in a given region S is said to be **meromorphic** in S. Thus a rational fractional function is meromorphic in the entire complex plane. The function $w = \tan z$ is meromorphic throughout the finite part of the plane, the poles being the zeros of $\cos z$.

As has been pointed out, a rational function may be either integral or fractional. We may now combine the results of Theorems I and III into one theorem for the unique characterization of rational functions. That theorem may be stated as follows:

THEOREM IV. *The necessary and sufficient condition that a single-valued analytic function is a rational function is that it has no essential singularities.*

We can now establish the following theorem.

THEOREM V. *A rational function is definitely determined, except for a constant factor, by its zero points and its poles in the finite portion of the plane.*

Let $f(z)$ be the given rational function. There can be but a finite number of zero points and a finite number of poles. Let the poles in the finite region be $\beta_1, \beta_2, \ldots, \beta_m$ having respectively the orders $k_1, k_2, \ldots, k_m$. These points must then appear as the zero points of the rational integral function $G_2(z)$ in the denominator of the given function. We may then write

$$G_2(z) = C_2 (z - \beta_1)^{k_1}(z - \beta_2)^{k_2} \ldots (z - \beta_m)^{k_m}.$$

The finite zero points of $f(z)$ appear as the zero points of the rational integral function $G_1(z)$ of $f(z)$. If we denote these zero points by $\alpha_1, \alpha_2, \ldots, \alpha_n$ and their orders respectively by $r_1, r_2, \ldots, r_n$, we have

$$G_1(z) = C_1 (z - \alpha_1)^{r_1}(z - \alpha_2)^{r_2} \ldots (z - \alpha_n)^{r_n}.$$

We have then

$$f(z) = \frac{G_1(z)}{G_2(z)} = C\frac{(z - \alpha_1)^{r_1}(z - \alpha_2)^{r_2} \ldots (z - \alpha_n)^{r_n}}{(z - \beta_1)^{k_1}(z - \beta_2)^{k_2} \ldots (z - \beta_m)^{k_m}}, \tag{4}$$

where $C = \dfrac{C_1}{C_2}$. In case $G_2(z)$ reduces to a constant, the function $f(z)$ is a rational integral function having but one pole, namely $z = \infty$.

Any rational function $f(z)$ can also be expressed as a constant or a rational integral function plus a finite number of rational fractions of the form

$$\frac{\alpha}{(z - \beta)^{\nu}},$$

where α, β are real or complex constants and ν is a positive integer. To express a given rational function $f(z)$ in this form consider the principal parts of the expansion at the various poles. Let the poles of $f(z)$ in the finite region be $\beta_1, \beta_2, \ldots, \beta_m$, having respectively the orders $k_1, k_2, \ldots, k_m$. Let the principal part of the expansion at β_1 be

$$\frac{c'_{-k_1}}{(z - \beta_1)^{k_1}} + \frac{c'_{-k_1+1}}{(z - \beta_1)^{k_1-1}} + \cdots + \frac{c'_{-1}}{(z - \beta_1)}.$$

Subtracting this from $f(z)$ we have left a function that is holomorphic in the entire finite portion of the plane except at $\beta_2, \ldots, \beta_m$. Proceeding in the same manner with the remaining poles just enumerated, we finally have a function that is holomorphic in the entire

finite plane and that has at most a pole at infinity; that is, we have a rational integral function, say of the form

$$c + c_1 z + c_2 z^2 + \cdots + c_n z^n.$$

We then have

$$\begin{aligned} f(z) = c &+ c_1 z + c_2 z^2 + \cdots + c_n z^n \\ &+ \frac{c'_{-k_1}}{(z-\beta_1)^{k_1}} + \frac{c'_{-k_1+1}}{(z-\beta_1)^{k_1-1}} + \cdots + \frac{c'_{-1}}{(z-\beta_1)} \\ &+ \frac{c''_{-k_2}}{(z-\beta_2)^{k_2}} + \frac{c''_{-k_2+1}}{(z-\beta_2)^{k_2-1}} + \cdots + \frac{c''_{-1}}{(z-\beta_2)} \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ &+ \frac{c^{(m)}_{-k_m}}{(z-\beta_m)^{k_m}} + \frac{c^{(m)}_{-k_m+1}}{(z-\beta_m)^{k_m-1}} + \cdots + \frac{c^{(m)}_{-1}}{(z-\beta_m)}. \end{aligned}$$

The terms in the first row of this expansion, aside from the constant term c, constitute the principal part of the expansion of $f(z)$ in the neighborhood of infinity. Consequently, we have the following theorem.

THEOREM VI. *A rational function is definitely determined except for an additive constant by the principal parts of its expansion at its various poles.*

The following examples furnish illustrations of Theorems V and VI and are left as exercises for the student.

Ex. 1. If a function has no other poles than simple poles at $z = 1, -1$, and its only zero points are of the first order and located at $z = i, -i$, show that it is necessarily of the form

$$f(z) = c\frac{z^2 + 1}{z^2 - 1}.$$

Ex. 2. Show that a function having no singularity either in the finite part of the plane or at infinity, other than a pole at the origin with the principal part

$$\frac{3}{z^3} + \frac{2}{z^2},$$

is of the form

$$f(z) = \frac{3}{z^3} + \frac{2}{z^2} + c = \frac{2z + 3}{z^3} + c.$$

THEOREM VII. *The sum of the residues of a rational function is zero.*

Let $f(z)$ be the given rational function. This function will have no essential singular points, but will have a finite number of poles, one of which may be at infinity. Let C be any circle in the finite region not inclosing any of the poles of $f(z)$ and not passing through

any such points. This circle, described in a clockwise direction, may be regarded as the boundary of the region exterior to it, that is the region containing all of the poles of $f(z)$.

The value of the integral

$$\frac{1}{2\pi i}\int_C f(z)\,dz,$$

taken in a clockwise direction, is by Theorem II, Art. 53, equal to the sum of the residues of $f(z)$. The value of the integral is unchanged except as to sign if it be taken in the counter-clockwise direction instead. But since $f(z)$ is holomorphic within and along C this latter integral vanishes. Hence taking the integral in the clockwise direction it is zero also. It follows then that the sum of the residues of $f(z)$ is necessarily zero as stated in the theorem.

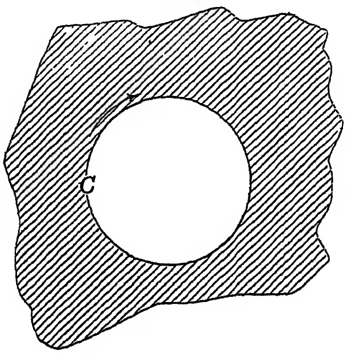

FIG. 100.

Some of the general properties of single-valued analytic functions that lead to special properties of rational functions may be readily deduced from a consideration of the **logarithmic derivative**, namely

$$\frac{f'(z)}{f(z)},$$

that is the quotient of the first derived function by the function itself. If the point $z = z_0$ is a regular point of $f(z)$ and if $f(z_0)$ is different from zero, then this point is a regular point of the logarithmic derivative. We shall see, however, that if $f(z)$ has a zero point at z_0, the logarithmic derivative is not holomorphic in the neighborhood of z_0. We have in fact the following theorem.

THEOREM VIII. *If $f(z)$ is holomorphic in the neighborhood of the point $z = z_0$ and has at this point a zero point of order k then the logarithmic derivative has at the same point a simple pole and a residue k.*

The given function $f(z)$ has a zero point at z_0 of the order k, and hence it can be written in the form

$$f(z) = (z - z_0)^k \phi(z), \tag{5}$$

where z_0 is a regular point of $\phi(z)$ and this function is different from zero for $z = z_0$. We have, therefore,

$$f'(z) = k\,(z - z_0)^{k-1}\phi(z) + (z - z_0)^k\phi'(z),$$

and hence obtain

$$\frac{f'(z)}{f(z)} = \frac{k}{z - z_0} + \frac{\phi'(z)}{\phi(z)}. \tag{6}$$

The function $\frac{\phi'(z)}{\phi(z)}$ is holomorphic in the neighborhood of z_0. The Laurent expansion of $\frac{f'(z)}{f(z)}$ in powers of $(z - z_0)$ contains but one term having a negative exponent, namely $\frac{k}{z - z_0}$; the rest of the terms have positive exponents, since they arise from the expansion of the holomorphic function $\frac{\phi'(z)}{\phi(z)}$. Hence, the residue is k and the point $z = z_0$ is a pole of order one, as required.

THEOREM IX. *If $f(z)$ has a pole of order k at $z = z_0$, then the logarithmic derivative has at the same point a simple pole and a residue $-k$.*

We may write

$$(z - z_0)^k f(z) = \phi(z),$$

where $\phi(z)$ is holomorphic in the neighborhood of the point $z = z_0$, and is different from zero for $z = z_0$.

We have then for $z \neq z_0$

$$f(z) = \frac{\phi(z)}{(z - z_0)^k},$$

and

$$f'(z) = \frac{\phi'(z)(z - z_0)^k - k(z - z_0)^{k-1}\phi(z)}{(z - z_0)^{2k}}.$$

Hence, we obtain

$$\frac{f'(z)}{f(z)} = \frac{\phi'(z)}{\phi(z)} - \frac{k}{z - z_0}. \tag{7}$$

Since $\frac{\phi'(z)}{\phi(z)}$ is holomorphic in the neighborhood of $z = z_0$, the expansion of the second member of the foregoing equation in powers of $(z - z_0)$ contains but one term with a negative exponent, namely $\frac{-k}{z - z_0}$, and hence the residue of the quotient $\frac{f'(z)}{f(z)}$ at z_0 is $-k$ and the point $z = z_0$ is a pole of order one, as stated in the theorem.

In the previous article we have shown that if $f(z)$ is holomorphic,

except for a finite number of poles, in a closed region bounded by a curve C, then the integral $\int_C f(z)\,dz$ is $2\pi i$ times the sum of the residues at the poles of that region. If we now apply this result in evaluating the integral of the logarithmic derivative, we obtain the following theorem.

THEOREM X. *The integral* $\frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}\,dz$ *taken in a positive sense around the boundary C of a closed region in which the rational function $f(z)$ is holomorphic except at a finite number of poles is equal to the number of zero points of $f(z)$ in this region diminished by the number of poles, each zero point and each pole being counted a number of times equal to its order.*

It should be noted that Theorems VIII, IX and X apply when z_0 is the point at infinity observing of course the convention as to the direction of integration about the point at infinity. In that case the given function may be written in the form

$$f(z) = z^{\lambda}\phi(z),$$

where $\phi(z)$ is holomorphic in the neighborhood of infinity and different from zero for $z = \infty$. If $\lambda = k > 0$, $f(z)$ has a pole of order k at infinity, and if $\lambda = -k$, $f(z)$ has a zero point of order k at infinity. We have then for the logarithmic derivative

$$\frac{f'(z)}{f(z)} = \frac{\lambda z^{\lambda-1}\phi(z) + z^{\lambda}\phi'(z)}{z^{\lambda}\phi(z)}$$
$$= \frac{\lambda}{z} + \frac{\phi'(z)}{\phi(z)}.$$

By Theorem III, Art. 53, the residue of $\frac{f'(z)}{f(z)}$ is $-\lambda$, provided $z^2\frac{\phi'(z)}{\phi(z)}$ is holomorphic in the neighborhood of $z = \infty$. This condition is easily seen to be satisfied by substituting $\frac{1}{z'}$ for z and examining the transformed function in the neighborhood of $z' = 0$.

This discussion leads to the following theorem.

THEOREM XI. *A rational function is just as often zero as infinite, when the entire complex plane including the point at infinity is considered, each zero point and each pole being counted a number of times equal to its order.*

This theorem follows at once from Theorem VII concerning the sum of the residues of a rational function. It was shown that this sum is necessarily zero. Hence, from Theorem X the number of zero points must equal the number of poles, each being counted a number of times equal to its order.

The foregoing theorem admits of a generalization as follows:

THEOREM XII. *A rational function of degree λ takes any given value, real or complex, exactly λ times.*

If $f(z)$ is rational, then $F(z) = f(z) - C$ is also a rational function, where C is any constant. By Theorem XI the function $F(z)$ must be as often zero as infinity. Hence, we may say that $f(z)$ takes any arbitrary value C as often as $F(z)$ becomes infinite. We shall now show that $F(z)$ becomes infinite a number of times equal to the degree of the function $f(z)$, thus establishing the theorem.

The degree of $F(z)$ is the same as that of $f(z)$. The function $F(z)$ may be written in the form

$$F(z) = \frac{\alpha_0 + \alpha_1 z + \cdots + \alpha_n z^n}{\beta_0 + \beta_1 z + \cdots + \beta_m z^m}, \qquad \alpha_n \neq 0, \qquad \beta_m \neq 0,$$

where the degree λ of $F(z)$ is the larger of the two numbers n, m, or equal to either m or n in case $m = n$. In any case $F(z)$ has a pole at every point where the denominator vanishes; for, by hypothesis the numerator and denominator have no common factor.

For $n = m = \lambda$ there are λ poles in the finite portion of the plane corresponding to the λ zero points of the denominator. We can show as follows that the point $z = \infty$ is a regular point of $F(z)$. Putting $z = \frac{1}{z'}$, the transformed function $\phi(z')$ is

$$\phi(z') = \frac{\alpha_0 + \frac{\alpha_1}{z'} + \cdots + \frac{\alpha_\lambda}{z'^\lambda}}{\beta_0 + \frac{\beta_1}{z'} + \cdots + \frac{\beta_\lambda}{z'^\lambda}} = \frac{\alpha_0 z'^\lambda + \alpha_1 z'^{\lambda-1} + \cdots + \alpha_\lambda}{\beta_0 z'^\lambda + \beta_1 z'^{\lambda-1} + \cdots + \beta_\lambda},$$

which is $\frac{\alpha_\lambda}{\beta_\lambda}$ for $z' = 0$. Since $\phi(z')$ is holomorphic in the neighborhood of the origin, the point $z = \infty$ is a regular point of $F(z)$. In this case then the number of poles of $F(z)$ is equal to the degree of the function.

For $n < m$, there are m poles of $F(z)$ in the finite region corresponding to the m zero points of the denominator. In this case also

the point $z = \infty$ is a regular point of $F(z)$, in fact $F(z)$ has a zero point at infinity; for, we have

$$\phi(z') = \frac{\alpha_0 + \frac{\alpha_1}{z'} + \cdots + \frac{\alpha_n}{z'^n}}{\beta_0 + \frac{\beta_1}{z'} + \cdots + \frac{\beta_m}{z'^m}} = \frac{\alpha_0 z'^m + \alpha_1 z'^{m-1} + \cdots + \alpha_n z'^{m-n}}{\beta_0 z'^m + \beta_1 z'^{m-1} + \cdots + \beta_m},$$

which is equal to zero for $z' = 0$. The total number of poles of $F(z)$ is therefore $m = \lambda$, the degree of $F(z)$.

If we have $n > m$, then in addition to the m poles corresponding to the zero points of the denominator, $F(z)$ has a pole of order $n - m$ at infinity; for, we have

$$\phi(z') = \frac{\alpha_0 + \frac{\alpha_1}{z'} + \cdots + \frac{\alpha_n}{z'^n}}{\beta_0 + \frac{\beta_1}{z'} + \cdots + \frac{\beta_m}{z'^m}} = \frac{\alpha_0 z'^n + \alpha_1 z'^{n-1} + \cdots + \alpha_n}{\beta_0 z'^n + \beta_1 z'^{n-1} + \cdots + \beta_m z'^{n-m}},$$

which has a pole of order $n - m$ at $z' = 0$. Hence $F(z)$ has a pole of the same order at infinity. In this case therefore the totality of poles is equal to $n = \lambda$, the degree of $F(z)$. Hence the theorem.

55. Transcendental functions. As we have seen, an analytic function that is not algebraic is a transcendental function. As all single-valued algebraic functions are necessarily rational, it follows that any single-valued analytic function that is not rational is transcendental, and hence we can now readily identify such functions by means of their singularities. A single-valued transcendental function must have an essential singularity; for, otherwise by Theorem IV, Art. 54, it would be a rational function. If a single-valued analytic function $G(z)$ has no singularity in the finite region and has an essential singularity at infinity, it is called a **transcendental integral function** of z. The expansion of such a function in a Maclaurin's series gives

$$G(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots + \alpha_n z^n + \cdots, \tag{1}$$

which converges for all finite values of z. If in (1) we replace z by $\frac{1}{z - z_0}$, we have a transcendental integral function of $\frac{1}{z - z_0}$, namely

$$G\left(\frac{1}{z - z_0}\right) = \alpha_0 + \frac{\alpha_1}{(z - z_0)} + \cdots + \frac{\alpha_n}{(z - z_0)^n} + \cdots. \tag{2}$$

From the form of the expansion it will be seen that this function has but one singular point, namely an essential singular point at $z = z_0$.

Conversely, if a single-valued analytic function has an essential singular point at $z = z_0$ and has no other singular points, then it can be expanded in the form given in (2) and hence is a transcendental integral function of $\dfrac{1}{z - z_0}$.

As we have seen, a power series may be integrated or differentiated term by term for values of the variable within the circle of convergence, and the resulting power series has the same circle of convergence as the given series. Consequently, the integral or the derived function of a transcendental integral function $G(z)$ is represented by a power series that converges for all finite values of z and hence is itself a transcendental integral function.

The following functions are transcendental integral functions:

$$e^z = 1 + z + \frac{z^2}{2\,!} + \frac{z^3}{3\,!} + \cdots,$$

$$\sin z = z - \frac{z^3}{3\,!} + \frac{z^5}{5\,!} - \frac{z^7}{7\,!} + \cdots,$$

$$\cos z = 1 - \frac{z^2}{2\,!} + \frac{z^4}{4\,!} - \frac{z^6}{6\,!} + \cdots,$$

$$\sinh z = z + \frac{z^3}{3\,!} + \frac{z^5}{5\,!} + \frac{z^7}{7\,!} + \cdots,$$

$$\cosh z = 1 + \frac{z^2}{2\,!} + \frac{z^4}{4\,!} + \frac{z^6}{6\,!} + \cdots.$$

A transcendental integral function differs from a rational integral function, in that it may have no zero points or it may have an infinite number of zero points. For example, the function

$$f(z) = e^z$$

is different from zero for all finite values of z. In fact we may state the following theorem.

THEOREM. *Any transcendental integral function $f(z)$ having no zero points may be written in the form*

$$f(z) = e^{G(z)},$$

where $G(z)$ is an integral function.

Since $f(z)$ is a transcendental integral function, $f'(z)$ is also a transcendental integral function, and as $f(z)$ has no zero points, then

$$F(z) = \frac{f'(z)}{f(z)}$$

is an integral function, as is also the function defined by the integral $\int_{z_0}^{z} F(z)\,dz$. We have then the integral function $\phi(z)$ defined by the relation

$$\phi(z) = \int_{z_0}^{z} F(z)\,dz = \int_{z_0}^{z} \frac{f'(z)}{f(z)}\,dz = \log f(z) - \log f(z_0).$$

Putting $\phi(z) + \log f(z_0) = G(z),$

we have $\log f(z) = G(z),$

whence $f(z) = e^{G(z)},$

where $G(z)$ is an integral function.

The function

$$f(z) = R(z)\,e^{G(z)},$$

where $R(z)$ is a rational integral function and $G(z)$ is an integral function, is a transcendental integral function having a finite number of zero points. As an example of a transcendental integral function having an infinite number of zero points, we have

$$f(z) = \sin z,$$

which is zero at the points

$$z = 0, \quad \pm\pi, \quad \pm 2\pi, \; . \; . \; . \; , \; \pm k\pi, \; . \; . \; . \; .$$

A transcendental integral function differs from a rational integral function in still another way. As the rational integral function has a pole at infinity, there always exists a circle about the origin such that for all points exterior to this circle we have $|f(z)| > M$, where M is an arbitrarily large number. Since a transcendental integral function has an essential singularity at $z = \infty$, the given function may be made to approach any value as z becomes infinite. Consequently, there are always values of z exterior to any circle however large about the origin for which $|f(z)| > M$ and also values of z for which $|f(z)| < \epsilon$, where ϵ is an arbitrarily small positive number.

A transcendental function having poles but no essential singular points in the finite portion of the plane is called a **transcendental fractional function.** This distinction between transcendental integral and transcendental fractional functions is suggested by the corresponding distinction between rational integral and rational fractional functions. In both cases a function is called integral when it has but one singular point and that is at $z = \infty$. Likewise in both cases we call a function fractional if it has poles and no other singular points in the finite region.

The functions

$$\frac{z}{\sin z}, \qquad \tan z, \qquad \sec z$$

are illustrations of transcendental fractional functions, for they have an essential singular point at infinity and poles but no other singular points in the finite region. Each may be written as the quotient of two integral functions. It will be shown later (Art. 57) that every transcendental fractional function can be written as the quotient of two integral functions.

We have pointed out that any rational function can be expressed as the sum of a rational integral function and fractions of the form

$$\frac{\alpha}{(z - z_0)^\nu}.$$

It is not difficult in that case to set up the function when the principal part of the expansion is known for each of the various singular points, which for rational functions consist of a finite number of poles. The corresponding problem for transcendental functions, namely, the problem of setting up a function with arbitrarily chosen singular points and with corresponding arbitrary principal parts, is much more difficult. The question of the existence of an analytic function having a given infinite set of singular points, with given principal parts, will be considered in the following article.

56. Mittag-Leffler's theorem. Suppose we have given any infinite set of numbers $z_1, z_2, \ldots, z_k, \ldots$, all different, having the property that

$$|z_1| \leqq |z_2| \leqq \cdots \leqq |z_k| \leqq \cdots,$$

and suppose that

$$\underset{k=\infty}{L} z_k = \infty.$$

Mittag-Leffler was the first to show* that there always exists a single-valued analytic function having these points and no others as singular points, with given principal parts of the form

$$G_k\left(\frac{1}{z - z_k}\right), \qquad k = 1, 2, 3, \ldots, \tag{1}$$

where G_k is an integral function of $\dfrac{1}{z - z_k}$, rational or transcendental. If, as in rational functions, we add together the functions (1) we have

* See *Encyklopädie d. Math. Wiss.*, Bd. II_2, p. 80.

in this case an infinite series. Whether the function defined by this series is everywhere holomorphic except at the points z_k depends upon the nature of the convergence of the series. Mittag-Leffler showed that by associating with each principal part $G_k\left(\frac{1}{z - z_k}\right)$ a suitably chosen polynomial the series can be made to converge uniformly and hence define a function having the desired properties. His theorem may be stated as follows:

THEOREM. *Given an infinite set of points*

$$z_1, z_2, z_3, \ldots, z_k, \ldots,$$

such that

$$0 < |z_1| \leqq |z_2| \leqq \cdots \leqq |z_k| \leqq \cdots, \quad \underset{k=\infty}{L}\, z_k = \infty.$$

There exists a single-valued analytic function which is holomorphic for all finite values of $z \neq z_k$ *and which has an arbitrarily chosen integral function of* $\frac{1}{z - z_k}$, *namely* $G_k\left(\frac{1}{z - z_k}\right)$, *as the principal part of its expansion in the neighborhood of* z_k.

Let
$$\sigma_1 + \sigma_2 + \cdots + \sigma_k + \cdots \tag{2}$$
be a convergent series of positive terms. The function

$$G_k\left(\frac{1}{z - z_k}\right), \qquad k = 1, 2, 3, \ldots,$$

is holomorphic everywhere in the finite region except for $z = z_k$. It can be expanded in a Maclaurin series, and this series converges and represents the function for all values of z within a circle C_k (Fig. 101) about the origin as a center and having $|z_k|$ as a radius. Consequently, we may write

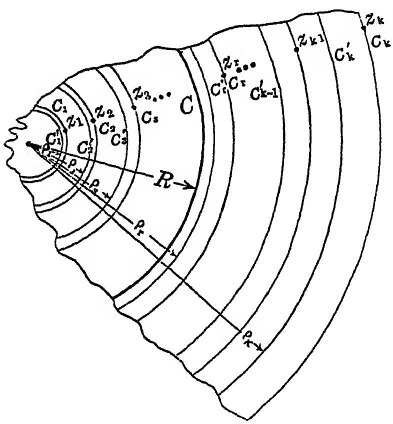

FIG. 101.

$$G_k\left(\frac{1}{z - z_k}\right) = \alpha_{0,k} + \alpha_{1,k}z + \alpha_{2,k}z^2 + \cdots + \alpha_{n,k}z^n + \cdots. \tag{3}$$

Within and upon a circle C_k' about the origin and having a radius $\rho_k = \theta\,|z_k|$, $0 < \theta < 1$, the series in the second member of (3) con-

verges uniformly. Then there exists an integer m_k such that for z within or upon C_k' we have

$$\left| \sum_{n=m_k-1}^{\infty} \alpha_{n,k} z^n \right| < \sigma_k; \tag{4}$$

that is, putting $P_k(z)$ equal to the polynomial

$$\alpha_{0,k} + \alpha_{1,k} z + \alpha_{2,k} z^2 + \cdot \cdot \cdot + \alpha_{m_k-2,k} z^{m_k-2}$$

formed by taking the first $m_k - 1$ terms of the series (3), we have, for z within or upon C_k',

$$\left| G_k\left(\frac{1}{z - z_k}\right) - P_k(z) \right| < \sigma_k. \tag{5}$$

Denote by R the radius of any circle C about the origin. Then in the sequence 1, 2, 3, . . . there can be found an integer, say r, such that we have

$$R < \theta \mid z_r \mid.$$

For all values of z within or upon the circle C, z must also lie within

$$C_r', \; C'_{r+1}, \; . \; . \; . \; ,$$

and hence for these values of z (4) holds for $k \geqq r$. For values of z within or upon the circle C consider the series

$$\sum_{k=r}^{\infty} \left\{ G_k\left(\frac{1}{z - z_k}\right) - P_k(z) \right\}. \tag{6}$$

By (5), each term of this series is less in absolute value than the corresponding term σ_k of the convergent series of positive terms $\sum_{k=r}^{\infty} \sigma_k$. Hence, by Weierstrass' test for uniform convergence (Theorem I, Art. 45) the series (6) converges absolutely and uniformly within and upon the circle C. As each term in this series is a holomorphic function, it follows that the series defines a function which is holomorphic in the region bounded by C.

The expression

$$\sum_{k=1}^{r-1} \left\{ G_k\left(\frac{1}{z - z_k}\right) - P_k(z) \right\} \tag{7}$$

is the sum of a finite number of functions, which sum is holomorphic in the region bounded by C, except for those points of the set z_1, z_2, . . . , z_{r-1} which lie within or upon C. Combining (6) and (7), we have a function

$$F_1(z) = \sum_{k=1}^{\infty} \left\{ G_k\left(\frac{1}{z - z_k}\right) - P_k(z) \right\}, \tag{8}$$

which is holomorphic in the region bounded by C, except for these same points. But since C is any circle about the origin it follows that the function $F_1(z)$ has the properties desired, and the theorem is established.

If now we add to the function $F_1(z)$ any integral function $G(z)$, rational or transcendental, we obtain a more general function

$$F(z) = F_1(z) + G(z),$$

which also has the same finite singular points with the same principal parts respectively as $F_1(z)$, and consequently $F(z)$ satisfies the conditions of the theorem. Conversely, if $F(z)$ is a function having the singular points

$$z_1, \ z_2, \ \ldots, \ z_k, \ \ldots$$

with the principal parts

$$G_1\left(\frac{1}{z - z_1}\right), \quad G_2\left(\frac{1}{z - z_2}\right), \ \ldots, \quad G_k\left(\frac{1}{z - z_k}\right), \ \ldots$$

respectively, then

$$F(z) - F_1(z)$$

is an integral function $G(z)$, and therefore we have

$$F(z) = F_1(z) + G(z).$$

For special cases a simpler form of the required function can be shown to exist. For example, let us consider the case where the function is to have simple poles at the points

$$z_1, \ z_2, \ \ldots, \ z_k, \ \ldots,$$

and where at each pole the residue is one. The principal part of the expansion in the neighborhood of each of the points z_k is then $\dfrac{1}{z - z_k}$. The series (3) then becomes

$$\frac{1}{z - z_k} = -\frac{1}{z_k} - \frac{z}{z_k^2} - \cdots - \frac{z^n}{z_k^{n+1}} - \cdots$$

and

$$P_k(z) = -\sum_{n=0}^{m_k-2} \frac{z^n}{z_k^{n+1}}.$$

The series (6) can then be written

$$\sum_{k=r}^{\infty} \left\{ G_k\left(\frac{1}{z - z_k}\right) - P_k(z) \right\} = -\sum_{k=r}^{\infty} \sum_{n=m_k-1}^{\infty} \frac{1}{z_k} \left(\frac{z}{z_k}\right)^n. \tag{9}$$

But we have $\left|\dfrac{z}{z_k}\right| < \theta$ for all values of z within or upon the circle C,

where $k \geqq r$. Therefore, for $k \geqq r$ and z within or upon C, we have

$$\sum_{n=m_k-1}^{\infty}\left|\frac{1}{z_k}\right|\cdot\left|\frac{z}{z_k}\right|^n = \frac{\left|\frac{1}{z_k}\right|\cdot\left|\frac{z}{z_k}\right|^{m_k-1}}{1-\left|\frac{z}{z_k}\right|} < \frac{1}{1-\theta}\frac{|z|^{m_k-1}}{|z_k|^{m_k}}.$$

Hence, from (9) we have

$$\sum_{k=r}^{\infty}\left|\left\{G_k\left(\frac{1}{z-z_k}\right)-P_k(z)\right\}\right| < \frac{1}{1-\theta}\sum_{k=r}^{\infty}\frac{|z|^{m_k-1}}{|z_k|^{m_k}} = \frac{1}{1-\theta}\cdot\frac{1}{z}\sum_{k=r}^{\infty}\left|\frac{z}{z_k}\right|^{m_k}. \tag{10}$$

It follows that the series of holomorphic functions in the left-hand member of (9) converges uniformly and represents a holomorphic function within and upon the circle C if the series $\sum_{k=r}^{\infty}\left|\frac{z}{z_k}\right|^{m_k}$ converges uniformly. By the Weierstrass test for uniform convergence this series converges uniformly if its terms are numerically less than the corresponding terms of a convergent series of positive terms. It is sufficient to take $m_k = k$; for, since

$$\left|\frac{z}{z_k}\right| < \theta < 1,$$

we have

$$\sum_{k=r}^{\infty}\left|\frac{z}{z_k}\right|^{m_k} < \sum_{k=r}^{\infty}\theta^k, \tag{11}$$

where $\sum_{k=r}^{\infty}\theta^k$ is a convergent series.

If there exists an integer p independent of k for which the series

$$\sum_{k=1}^{\infty}\left|\frac{1}{z_k}\right|^p \tag{12}$$

converges, then the series

$$\sum_{k=r}^{\infty}\left|\frac{z}{z_k}\right|^p$$

also converges and we may put $m_k = p$. In this case, therefore, the degree of the polynomial $P_k(z)$ does not need to increase indefinitely with k.

Consequently, if

$$z_1, z_2, \ldots, z_k, \ldots$$

are to be simple poles and if the function $F_1(z)$ has the residue 1 at each pole, then from (8) $F_1(z)$ takes the simple form

$$F_1(z) = \sum_{k=1}^{\infty}\left\{\frac{1}{z-z_k}+\frac{1}{z_k}+\frac{z}{z_k^2}+\cdots+\frac{z^{\nu-2}}{z_k^{\nu-1}}\right\}. \tag{13}$$

If there exists a constant p as given by (12), then we may put $\nu = p$. In case no such constant exists, then we need at most put * $\nu = k$.

In Art. 44 it was shown that the series

$$\sum{}' \left|\frac{1}{\Omega}\right|^3$$

converges. By aid of the special case just considered, it follows that the function

$$\zeta(z) = \frac{1}{z} + \sum{}' \left\{ \frac{1}{z - \Omega} + \frac{1}{\Omega} + \frac{z}{\Omega^2} \right\} \tag{14}$$

has simple poles of residue one at $z = 0$ and at each of the points $z = \Omega$. With the exception of these points the function is holomorphic in the entire finite plane.

Differentiating (14) term by term and changing the sign of each term we have another important function namely

$$\wp(z) = -\zeta'(z) = \frac{1}{z^2} + \sum{}' \left\{ \frac{1}{(z - \Omega)^2} - \frac{1}{\Omega^2} \right\}. \tag{15}$$

By differentiating again, we have

$$\wp'(z) = -\frac{2}{z^3} - 2\sum{}' \frac{1}{(z - \Omega)^3}. \tag{16}$$

The three functions (14), (15), (16) are made use of by Weierstrass in the theory of elliptic functions.

57. Expansion of functions in infinite products. In addition to expanding an analytic function by means of an infinite series, another method is often employed, namely infinite products. Suppose we have the infinite sequence

$$(1 + \alpha_1),\ (1 + \alpha_2),\ \ldots,\ (1 + \alpha_k),\ \ldots.$$

The continued product of the first n elements may be denoted by

$$\Pi_n \equiv \prod_{k=1}^{n} (1 + \alpha_k). \tag{1}$$

If at most a finite number of the factors $(1 + \alpha_k)$ are zero, that is if there exists a positive integer m_1 such that for $k \geqq m_1$ all of the factors are different from zero, then the product (1) is said to converge if the limit

$$\underset{n=\infty}{L} \prod_{k=m_1}^{n} (1 + \alpha_k),$$

* Borel has shown that it is sufficient to put $\nu > \log k$ (*Leçons sur les fonctions entières*, p. 10).

exists and is different from zero. If the product tends toward zero, or becomes infinite, or if for any reason it does not have a limit as n increases indefinitely, then the infinite product is called divergent.

The necessary and sufficient condition that an infinite product converges may be stated as follows:

THEOREM I. *The necessary and sufficient condition that an infinite product converges is that corresponding to an arbitrary positive number* ϵ *there exists a positive integer* m *such that for all values of* $n > m$ *we have*

$$\left|\prod_{k=n+1}^{n+p}(1+\alpha_k)-1\right|<\epsilon, \qquad p=1,\,2,\,\ldots. \tag{2}$$

Consider the sequence of the products

$$\prod_{m_1}^{m_1+1},\quad \prod_{m_1}^{m_1+2},\quad \ldots,\ \prod_{m_1}^{n},\ \ldots,\quad n>m_1. \tag{3}$$

By Theorem VI, Art. 12, the necessary and sufficient condition that this sequence converges is that for every positive number ϵ_1 there exists an integer m such that

$$\left|\prod_{m_1}^{n+p}-\prod_{m_1}^{n}\right|<\epsilon_1,\qquad n>m>m_1,\qquad p=1,\,2,\,3,\,\ldots. \tag{4}$$

We shall show that this condition is equivalent to the one given in the foregoing theorem. If the given infinite product converges, then we have

$$\underset{n=\infty}{L}\prod_{m_1}^{n}=A\neq 0.$$

There then exists a number $M>0$, such that for all values of $n>m_1$, we have

$$\left|\prod_{m_1}^{n}\right|>M. \tag{5}$$

Dividing (4) by (5) we get

$$\left|\frac{\prod_{m_1}^{n+p}}{\prod_{m_1}^{n}}-1\right|<\frac{\epsilon_1}{M},\qquad n>m,\qquad p=1,2,3,\,\ldots.$$

Putting $\frac{\epsilon_1}{M} = \epsilon$, this result may be written

$$\left| \prod_{n+1}^{n+p} - 1 \right| < \epsilon, \qquad n > m, \qquad p = 1, 2, 3, \ldots,$$

which is the condition given in the theorem.

Conversely, suppose we have given the inequality (2). We may write

$$\left| \frac{\left|\prod_{m_1}^{n+p}\right|}{\left|\prod_{m_1}^{n}\right|} - 1 \right| \leqq \left| \frac{\prod_{m_1}^{n+p}}{\prod_{m_1}^{n}} - 1 \right|.$$

But from the condition (2), we have by aid of the above relation

$$\left| \frac{\left|\prod_{m_1}^{n+p}\right|}{\left|\prod_{m_1}^{n}\right|} - 1 \right| < \epsilon, \qquad n > m, \qquad p = 1, 2, 3, \ldots; \tag{6}$$

that is, we have

$$-\epsilon < \frac{\left|\prod_{m_1}^{n+p}\right|}{\left|\prod_{m_1}^{n}\right|} - 1 < \epsilon,$$

or

$$(1-\epsilon)\left|\prod_{m_1}^{n}\right| < \left|\prod_{m_1}^{n+p}\right| < (1+\epsilon)\left|\prod_{m_1}^{n}\right|, \quad n > m, \quad p = 1, 2, 3, \ldots. \tag{7}$$

Suppose we now give to n any definite value greater than m, say $n = m_1 + \nu$. Since ϵ may be chosen arbitrarily, it follows from (7) that each element $\prod_{m_1}^{n}$ of the sequence (3) is in absolute value less than a positive number N, which is the largest number in the finite sequence

$$\left|\prod_{m_1}^{m_1+1}\right|, \quad \left|\prod_{m_1}^{m_1+2}\right|, \ldots, \left|\prod_{m_1}^{m_1+\nu-1}\right|, \quad (1+\rho)\left|\prod_{m_1}^{m_1+\nu}\right|,$$

where ρ is any constant greater than zero. Likewise each element of (3) is larger in absolute value than a positive number N_1, which is the smallest number in the finite sequence

$$\left|\prod_{m_1}^{m_1+1}\right|, \quad \left|\prod_{m_1}^{m_1+2}\right|, \ldots, \left|\prod_{m_1}^{m_1+\nu-1}\right|, \quad (1-\rho)\left|\prod_{m_1}^{m_1+\nu}\right|.$$

From (6) we have

$$\left|\prod_{m_1}^{n+p} - \prod_{m_1}^{n}\right| < \epsilon \left|\prod_{m_1}^{n}\right|, \quad n > m, \quad p = 1, 2, 3, \ldots .$$

But since we have $\left|\prod_{m_1}^{n}\right| < N$, it follows that by putting

$$\epsilon = \frac{\epsilon'}{N},$$

where ϵ' is arbitrarily small, we have

$$\left|\prod_{m_1}^{n+p} - \prod_{m_1}^{n}\right| < \epsilon', \qquad p = 1, 2, 3, \ldots .$$

Hence, the condition given in (4) is satisfied and the limit

$$\underset{n=\infty}{L} \prod_{m_1}^{n}$$

exists. Suppose the limit is A. Since $\left|\prod_{m_1}^{n}\right|$ is always greater than the positive number N_1, it follows that $A \neq 0$. Hence the given sequence converges and the demonstration of the theorem is completed.

If the product $\prod (1 + |\alpha_k|)$ converges, then the product $\prod (1+\alpha_k)$ is said to **converge absolutely.** If an infinite product converges but does not converge absolutely, it is said to **converge conditionally.** As a condition for absolute convergence, we have the following theorem.

THEOREM II. *If the series* $\sum_{k=1}^{\infty} \alpha_k$ *converges absolutely, then the infinite product*

$$\prod_{k=1}^{\infty} (1 + \alpha_k)$$

converges absolutely.

From the convergence of the series $\Sigma\alpha_k$, it follows that only a finite number of the factors $(1 + \alpha_k)$ can be zero. We assume as before that for $k \geqq m_1$ the factors are all different from zero. We may then write for $n > m_1$

$$\prod_{k=m_1}^{n} (1 + \alpha_k) = e^{\log(1+\alpha_{m_1})+\log(1+\alpha_{m_1+1}) + \cdots + \log(1+\alpha_k) + \cdots + \log(1+\alpha_n)}. \quad (8)$$

The given product will then converge absolutely if the series

$$\log(1+|\alpha_{m_1}|)+\log(1+|\alpha_{m_1+1}|)+\cdots+\log(1+|\alpha_k|)+\cdots \quad (9)$$

converges. But this series converges, if the series

$$|\alpha_1|+|\alpha_2|+\cdots+|\alpha_k|+\cdots \quad (10)$$

converges;* for, the ratio of the general terms of the two series, namely

$$\frac{\log(1+|\alpha_k|)}{|\alpha_k|},$$

has the limit 1 as k becomes infinite. Since the series (10) converges by hypothesis, that is $\Sigma\alpha_k$ converges absolutely, the theorem follows.

In the discussion of rational integral functions we saw that such functions are determined except as to a constant factor when the zero points are known. In the case of transcendental integral functions we saw that the given function might have no zero points, or on the other hand it might have an infinite number of such points. If we have given an infinite set of points

$$z_1,\ z_2,\ z_3,\ \ldots,\ z_k,\ \ldots,$$

having $z = \infty$ as a limiting point, it is of interest to see whether an integral function can be set up having these points and no others as zero points. Clearly such a function must be a transcendental function, since a rational integral function can have but a finite number of zero points. Weierstrass has shown how the desired function can be represented by an infinite product. It is at once clear that if $\Phi(z)$ is such a function and $G(z)$ denotes an integral function, then

$$F(z) = \Phi(z)e^{G(z)}$$

is also such a function, since as we have seen $e^{G(z)}$ can have no zero points. The function $e^{G(z)}$ plays the same rôle that the constant factor does in the representation of a rational integral function as the product of a finite number of binomials.

We may now state the following theorem.

THEOREM III. *Given a set of points*

$$z_1,\ z_2,\ \ldots,\ z_k,\ \ldots \quad (11)$$

not including the origin such that

$$|z_1| \leqq |z_2| \leqq \cdots \leqq |z_k| \leqq \cdots,\ \underset{k=\infty}{L}\, z_k = \infty.$$

* See Bromwich, *Theory of Infinite Series*, Art. 9.

There exists a transcendental integral function of the form

$$\Phi(z) = \prod_{k=1}^{\infty}\left(1 - \frac{z}{z_k}\right)e^{\frac{z}{z_k}+\frac{1}{2}\left(\frac{z}{z_k}\right)^2+\cdots+\frac{1}{m_k-1}\left(\frac{z}{z_k}\right)^{m_k-1}}$$

having the points z_k and no others as zero points. Moreover, the function

$$F(z) = e^{G(z)}\cdot\Phi(z)$$

is the most general function having this property.

Consider the infinite product

$$\prod_{k=1}^{\infty}\left(1 - \frac{z}{z_k}\right)e^{\phi_k\left(\frac{z}{z_k}\right)}, \tag{12}$$

where $\phi_k\left(\dfrac{z}{z_k}\right)$ is a rational integral function of $\left(\dfrac{z}{z_k}\right)$ as yet undetermined. The factor

$$\left(1 - \frac{z}{z_k}\right)e^{\phi_k\left(\frac{z}{z_k}\right)}$$

is called a **primary factor.** This factor has one and only one zero, namely $z = z_k$. We shall show how the functions $\phi_k\left(\dfrac{z}{z_k}\right)$, $k = 1, 2, \ldots,$ can be determined so that the infinite product (12) will converge in an arbitrarily large circle and define a function having the required properties.

Let C_k (Fig. 101) be the circle about the origin as a center having $|z_k|$ as a radius. Let C_k' be a circle concentric with C_k and having the radius

$$\rho_k = \theta\,|z_k|, \qquad 0 < \theta < 1.$$

Denote by R the radius of any circle C about the origin. Then there exists an integer r such that C lies within

$$C_r', \; C'_{r+1}, \; \ldots,$$

and that within and upon C we have at most the points

$$z_1, \; z_2, \; \ldots \; z_{r-1}. \tag{13}$$

The product of the corresponding primary factors is an integral function having the points (13) as zero points. The remaining factors of (12) give rise to the product

$$\prod_{k=r}^{\infty}\left(1 - \frac{z}{z_k}\right)e^{\phi_k\left(\frac{z}{z_k}\right)}. \tag{14}$$

For any integer q we have

$$\prod_{k=r}^{r+q}\left(1-\frac{z}{z_k}\right)e^{\phi_k\left(\frac{z}{z_k}\right)} = e^{\sum_{k=r}^{r+q}\left\{\log\left(1-\frac{z}{z_k}\right)+\phi_k\left(\frac{z}{z_k}\right)\right\}}. \tag{15}$$

Hence, the convergence of the infinite product (14) will follow from the convergence of the series

$$\sum_{k=r}^{\infty}\left\{\log\left(1-\frac{z}{z_k}\right)+\phi_k\left(\frac{z}{z_k}\right)\right\}. \tag{16}$$

Since all of the points z_k, $k \geqq r$, lie outside of the circle C, it follows that if the right-hand member of (15) converges, it defines a function that does not vanish for z within or upon C. Consequently, the same may be said of the infinite product (14), provided this product is convergent.

Expanding $\log\left(1-\frac{z}{z_k}\right)$ in a power series, we have

$$\log\left(1-\frac{z}{z_k}\right) = -\frac{z}{z_k}-\frac{1}{2}\left(\frac{z}{z_k}\right)^2-\frac{1}{3}\left(\frac{z}{z_k}\right)^3-\cdots-\frac{1}{n}\left(\frac{z}{z_k}\right)^n-\cdots. \tag{17}$$

Suppose we let $\phi_k\left(\frac{z}{z_k}\right)$ be of degree m_k-1, where m_k-1 is to be determined later. We may then put

$$\phi_k\left(\frac{z}{z_k}\right) = \frac{z}{z_k}+\frac{1}{2}\left(\frac{z}{z_k}\right)^2+\cdots+\frac{1}{m_k-1}\left(\frac{z}{z_k}\right)^{m_k-1}.$$

From (16) we now have

$$\sum_{k=r}^{\infty}\left\{\log\left(1-\frac{z}{z_k}\right)+\phi_k\left(\frac{z}{z_k}\right)\right\} = -\sum_{k=r}^{\infty}\sum_{n=m_k}^{\infty}\frac{1}{n}\left(\frac{z}{z_k}\right)^n. \tag{18}$$

We have

$$R < \theta\,|z_k|, \qquad k \geqq r.$$

For all values of z within or upon the circle C, we have then

$$\left|\frac{z}{z_k}\right| < \theta, \qquad k \geqq r.$$

By aid of this relation we obtain

$$\sum_{n=m_k}^{\infty}\frac{1}{n}\left|\frac{z}{z_k}\right|^n < \sum_{n=m_k}^{\infty}\left|\frac{z}{z_k}\right|^n = \frac{\left|\frac{z}{z_k}\right|^{m_k}}{1-\left|\frac{z}{z_k}\right|} < \frac{1}{1-\theta}\left|\frac{z}{z_k}\right|^{m_k}.$$

Hence, from (18) we have

$$\left|-\sum_{k=r}^{\infty}\sum_{n=m_k}^{\infty}\frac{1}{n}\left(\frac{z}{z_k}\right)^n\right| < \frac{1}{1-\theta}\sum_{k=r}^{\infty}\left|\frac{z}{z_k}\right|^{m_k}. \tag{19}$$

It follows then that the series (14) converges uniformly if the series

$$\sum_{k=r}^{\infty}\left|\frac{z}{z_k}\right|^{m_k}$$

converges uniformly, which it does if $m_k = k$, as we have seen (Art. 56).

Each term of the series (16) is holomorphic for values of z within and upon the circle C, and since the convergence is uniform, it follows that (16) and hence (14) defines a function which is holomorphic within and upon C. But C is any circle about the origin as a center, hence for m_k equal to k the product

$$\prod_{k=1}^{\infty}\left(1-\frac{z}{z_k}\right)e^{\frac{z}{z_k}+\cdots+\frac{1}{m_k-1}\left(\frac{z}{z_k}\right)^{m_k-1}}$$

defines a transcendental integral function $\Phi(z)$ having the required properties. If we desire the most general function of this type, all we need to do is to introduce as a factor the most general function that has no roots, namely the function $e^{G(z)}$, where $G(z)$ is an integral function.

In the foregoing discussion the origin was not included in the set of zero points. If it is desired to include that point as a zero point, say of the order λ, all that is needed is to add the factor z^λ and write the function

$$F(z) = e^{G(z)}\,z^\lambda\,\Phi(z), \tag{20}$$

which is also a transcendental integral function. It is evident that by varying the function $G(z)$ in (20) we may obtain an infinite number of transcendental integral functions having the points given in (11) as zero points.

We have seen that there may exist an integer p independent of k which causes the series

$$\sum\left|\frac{1}{z_k}\right|^p$$

to converge and hence also the infinite product defining $\Phi(z)$. We may then put $m_k = p$ and have

$$F(z) = e^{G(z)}z^\lambda\prod_{k=1}^{\infty}\left(1-\frac{z}{z_k}\right)e^{\frac{z}{z_k}+\frac{1}{2}\left(\frac{z}{z_k}\right)^2+\cdots+\frac{1}{p-1}\left(\frac{z}{z_k}\right)^{p-1}}. \tag{21}$$

In the discussion of integral functions it is desirable to introduce what is known as the class of the function. For this purpose let us

suppose that p is the smallest integer that makes the series $\sum \left|\frac{1}{z_k}\right|^p$ converge. Let us also suppose $G(z)$ to be a polynomial, say of degree g. Then the **class*** of the integral function $F(z)$ given by (21) is defined as the larger of the two integers $p - 1$ and g. Since the degree of $G(z)$ can be changed without affecting the zero points of $F(z)$, we may so choose the polynomial $G(z)$ that g is less than $p - 1$ and hence $p - 1$ is then the class of $F(z)$. The class of any rational integral function is zero, as is also that of the transcendental integral function

$$f(z) = \prod_{k=1}^{\infty}\left(1 - \frac{z}{k^2}\right).$$

On the other hand, the function

$$\sin z = z \prod_{k=-\infty}^{\infty}\left(1 - \frac{z}{k\pi}\right)e^{\frac{z}{k\pi}}$$

is of class one.

We have seen, Art. 44, for example, that the series

$$\sum{}' \left|\frac{1}{\Omega}\right|^3$$

converges. Hence there exists a transcendental integral function of the form

$$z\prod{}'\left(1 - \frac{z}{\Omega}\right)e^{\frac{z}{\Omega} + \frac{1}{2}\left(\frac{z}{\Omega}\right)^2} = z \cdot e^{\sum'\left\{\log\left(1-\frac{z}{\Omega}\right) + \frac{z}{\Omega} + \frac{1}{2}\left(\frac{z}{\Omega}\right)^2\right\}}$$

having the Ω points and also the point $z = 0$ and no others as zero points. This function is the σ-function of Weierstrass and is of great importance in his development of the theory of elliptic functions. It is evidently a transcendental integral function of class two.

It has been seen that every rational fractional function is the quotient of two rational integral functions. Such a function is uniquely characterized by the fact that it has at most a pole at $z = \infty$ and only a finite number of poles in the finite region of the complex plane. In a similar manner, we have defined a transcendental fractional function as one that has an essential singularity at $z = \infty$ and only poles in the finite region. These poles may, however, be dense at the point $z = \infty$. We can now demonstrate the following theorem.

* The term *class* is here used as the equivalent of the French word *genre* introduced by Laguerre, and the German word Höhe, introduced by V. Schaper. See Borel, *Leçons sur les fonctions entières*, p. 25; also Osgood, *Encyklopadie d. Math. Wiss.*, Vol. II_2, Part I, p. 79.

THEOREM IV. *Every transcendental fractional function can be expressed as the quotient of two integral functions.*

Let the points

$$z_1, z_2, \ldots, z_k, \ldots, \tag{22}$$

where

$$|z_1| \leqq |z_2| \leqq \ldots \leqq |z_k| \leqq \ldots, \underset{k=\infty}{L} z_k = \infty,$$

be the poles of the given transcendental function $f(z)$. If any of the poles is of an order higher than one, we shall regard a number of the points z_k equal to the order of the pole as coincident. By Theorem III there exists a transcendental integral function $G_2(z)$ having the points (22) and no others as zero points. Moreover, at each point the order of the zero of $G_2(z)$ is the same as the order of the pole of $f(z)$; for, in each case the number of coincident points z_k is the same. It follows that the product $G_2(z) \cdot f(z)$ has no singular points in the finite region. Consequently, we have

$$G_2(z) f(z) = G_1(z),$$

where $G_1(z)$ is an integral function. It follows then that

$$f(z) = \frac{G_1(z)}{G_2(z)},$$

as the theorem requires.

58. Periodic functions. In the discussion of elementary functions in Chapter IV, attention was called to the fact that certain of those functions are simply periodic; that is, the function remains invariant under a translation of the plane by means of the relation

$$z' = z + n\omega,$$

where n is an integer and ω is the primitive period of the function. If $f(z)$ is the given function, we then have

$$f(z + n\omega) = f(z). \tag{1}$$

In the illustrations considered the period-strips were taken parallel to the axes of coördinates. It is not necessary, however, to choose the strips in that manner; for, if we locate the points

$$z_0 + n\omega,$$

where z_0 is any point, and draw parallel lines through these points making a convenient angle different from zero or $\frac{\pi}{2}$ with the X-axis, the strips bounded by these lines may be taken as period-strips. As

a matter of fact, the boundary lines of the regions of periodicity need not even be straight lines; for, all that is essential is that the plane be divided into congruent strips such that for any point z in any strip there corresponds a point $z + n\omega$ in each of the other strips for which equation (1) holds. It is of importance also to note that a given function may repeat its values in a period-strip; for, as we have seen in the case of $w = \cos z$, the period-strips are not identical with the fundamental regions of the function.

Functions like the exponential function and the trigonometric functions are, as we have seen, simply periodic. Single-valued analytic functions may, however, have two independent periods, where we understand two periods of a function to be independent if they are not integral multiples of the same primitive period. For example, a function is said to be **doubly periodic** if it has two periods $2\,\omega_1$, $2\,\omega_3$, which are independent of each other and of z, such that a translation of the plane by either of the relations

$$z' = z + 2\,\omega_1, \tag{2}$$

$$z'' = z + 2\,\omega_3 \tag{3}$$

leaves the function unchanged. We have then the two relations

$$f(z + 2\,\omega_1) = f(z), \tag{3}$$

$$f(z + 2\,\omega_3) = f(z). \tag{4}$$

As in the case of simply periodic functions, any translation of the complex plane by means of the relations

$$z' = z + 2\,m_1\omega_1, \qquad m_1 = \pm 1, \quad \pm 2, \; . \; . \; . \; , \tag{5}$$

$$z'' = z + 2\,m_3\omega_3, \qquad m_3 = \pm 1, \quad \pm 2, \; . \; . \; . \; , \tag{6}$$

also leaves the function unchanged. It follows at once that any combination of the translations (5), (6) leaves the given function invariant, since any such combination may be regarded as a succession of translations by means of (2), (3). We may then write

$$f(z + 2\,m_1\omega_1 + 2\,m_3\omega_3) = f(z), \tag{7}$$

which shows that

$$\Omega \equiv 2\,m_1\omega_1 + 2\,m_3\omega_3 \tag{8}$$

is likewise a period of the given function.

If all of the periods of the given function can be written in the form (8), that is if every such period can be expressed as the sum of integral multiples of these two periods, then $2\,\omega_1$, $2\,\omega_3$ are called a **primitive period pair.**

In order that any two pairs other than $2\,\omega_1$, $2\,\omega_3$, for example

$$2\,m_1'\omega_1 + 2\,m_3'\omega_3,$$
$$2\,m_1''\omega_1 + 2\,m_3''\omega_3,$$

may be taken as a primitive pair, it is sufficient that we have

$$\Delta \equiv \begin{vmatrix} m_1' & m_3' \\ m_1'' & m_3'' \end{vmatrix} = \pm 1. \tag{9}$$

For, let $2\,\omega_1'$, $2\,\omega_3'$ be any two independent periods of $f(z)$ other than $2\,\omega_1$, $2\,\omega_3$. Putting

$$2\,m_1'\,\omega_1 + 2\,m_3'\omega_3 = 2\,\omega_1',$$
$$2\,m_1''\,\omega_1 + 2\,m_3''\omega_3 = 2\,\omega_3',$$

and solving for $2\,\omega_1$, $2\,\omega_3$, we have

$$2\,\omega_1 = \frac{2\,m_3''\omega_1' - 2\,m_3'\omega_3'}{\Delta}, \qquad 2\,\omega_3 = \frac{-2\,m_1''\omega_1' + 2\,m_1'\omega_3'}{\Delta}. \tag{10}$$

If $\Delta = \pm 1$, then $2\,\omega_1$, $2\,\omega_3$ are each a sum of multiples of $2\,\omega_1'$, $2\,\omega_3'$ and consequently any period

$$\Omega = 2\,m_1\omega_1 + 2\,m_3\omega_3$$

of $f(z)$ can be written in the form

$$\Omega = 2\,n_1\omega_1' + 2\,n_3\omega_3';$$

hence $2\,\omega_1'$, $2\,\omega_3'$ are a primitive period pair.

THEOREM I. *Let $f(z)$ be a single-valued doubly periodic analytic function with the independent periods $2\,\omega_1$, $2\,\omega_3$. Then if $f(z)$ is not a constant, the ratio $\frac{\omega_3}{\omega_1}$ can not be real.*

Let us assume first of all that the ratio $\frac{\omega_3}{\omega_1}$ is real and commensurable, say

$$\frac{\omega_3}{\omega_1} = \frac{p}{q},$$

where p, q are integers prime to each other. We shall show that this assumption leads to a conclusion which is contrary to the given hypothesis. Since $2\,\omega_1$, $2\,\omega_3$ are periods, it follows that $2\,m_1\omega_1 + 2\,m_3\omega_3$ is also a period. But from the assumed relation, we may write

$$2\,m_1\omega_1 + 2\,m_3\omega_3 = 2\,\omega_1\left(m_1 + m_3\frac{p}{q}\right) = 2\,\omega_3\left(m_1\frac{q}{p} + m_3\right)$$
$$= \frac{2\,\omega_1(m_1 q + m_3 p)}{q} = \frac{2\,\omega_3(m_1 q + m_3 p)}{p}. \tag{11}$$

Since p, q are relatively prime, m_1 and m_3 can be so chosen that *

$$m_1 q + m_3 p = 1.$$

Consequently, we obtain from (11)

$$2\,m_1\omega_1 + 2\,m_3\omega_3 = \frac{2\,\omega_1}{q} = \frac{2\,\omega_3}{p} = 2\,\omega,$$

where $2\,\omega$ is a period. Hence, we have

$$2\,\omega_1 = 2\,q\omega, \qquad 2\,\omega_3 = 2\,p\omega;$$

that is, the periods $2\,\omega_1$, $2\,\omega_3$ are each multiples of a common period $2\,\omega$ and hence are not independent.

Likewise the assumption that the ratio $\frac{\omega_3}{\omega_1}$ is real and incommensurable leads to a contradiction. For, let $\frac{\omega_3}{\omega_1}$ be converted into a continued fraction. Then $\frac{\omega_3}{\omega_1}$ must lie between any two consecutive convergents,† say

$$\frac{p_k}{q_k}, \qquad \frac{p_{k+1}}{q_{k+1}}.$$

Hence, the value of $\frac{\omega_3}{\omega_1}$ differs from either of these convergents by less than $\frac{1}{q_k q_{k+1}}\cdot$ We may then write

$$\left|\frac{\omega_3}{\omega_1} - \frac{p_k}{q_k}\right| < \frac{1}{q_k q_{k+1}},$$

whence

$$|\,q_k\omega_3 - p_k\omega_1\,| < \frac{|\,\omega_1\,|}{q_{k+1}}.$$

But since q_{k+1} can be taken as large as we please, it follows that $\frac{|\,\omega_1\,|}{q_{k+1}}$ may be made as small as we please. From the foregoing relation, it follows that however small $\frac{|\,\omega_1\,|}{q_{k+1}}$ may be, values of m_1 and m_3 exist such that $|\,2\,m_1\omega_1 + 2\,m_3\omega_3\,| \equiv |\,\Omega\,|$ is numerically arbitrarily small. Consequently, there exists a set of values z dense at any regular point z_0 for which $f(z)$ has the same value $f(z_0)$. This condition can not exist except when $f(z)$ is a constant.

Since the ratio $\frac{\omega_3}{\omega_1}$ can be neither real and commensurable nor real and incommensurable it must be complex.

* See Chrystal, *Text-Book of Algebra*, Vol. II, p. 409.

† *Ibid.*, p. 410.

Let $2\,\omega_1$, $2\,\omega_3$ be a primitive period pair of the single-valued doubly periodic analytic function $f(z)$. Since the ratio $\frac{\omega_3}{\omega_1}$ can not be real, the straight line joining the origin with $2\,\omega_1$ makes an angle different from zero or π with the line joining the origin with $2\,\omega_3$. Consequently, the set of complex values

$$\Omega = 2\,m_1\omega_1 + 2\,m_3\omega_3$$

is represented by a **net of points** covering the complex plane. Moreover, any other primitive period pair of $f(z)$ must lead to the same net. If z_0 is any point in the region of existence of $f(z)$, then in the parallelogram

$$z_0,\quad z_0 + 2\,\omega_1,\quad z_0 + 2\,\omega_3,\quad z_0 + 2\,\omega_1 + 2\,\omega_3,$$

the given function takes all of its values. This parallelogram is called a **primitive period-parallelogram.** If we put

$$2\,\omega_1 + 2\,\omega_3 = -2\,\omega_2,$$

we have the relation

$$2\,\omega_1 + 2\,\omega_2 + 2\,\omega_3 = 0.$$

By drawing parallel straight lines through the points of the net $z_0 + \Omega$ as shown in Fig. 102, we have a set of congruent period-

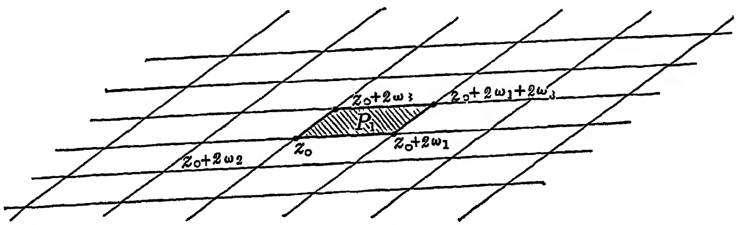

FIG. 102.

parallelograms covering the entire complex plane. If z_0 is a singular point, then each point $z_0 + \Omega$ is likewise a singular point. It is often convenient to choose z_0 so that no singular point of $f(z)$ lies upon the boundary of the period-parallelograms. This choice is always possible if the number of singular points of $f(z)$ in a period-parallelogram is finite. For example, if the singular points of $f(z)$ are restricted to poles, then there are but a finite number of poles in the initial period-parallelogram, and hence by the proper choice of z_0 none of these poles will lie upon the boundary of any period-parallelogram.

We shall now set up the convention that the points on one pair of adjacent boundary lines of a primitive period-parallelogram belong to the parallelogram and that the points on the other two boundary lines do not. For example, in the parallelogram P_1, Fig. 102, the points on the boundary joining z_0, $z_0 + 2\,\omega_1$ and z_0, $z_0 + 2\,\omega_3$, the points $z_0 + 2\,\omega_1$, $z_0 + 2\,\omega_3$ excepted, are considered as belonging to the period-parallelogram but the points on the other two boundary lines do not. It is sufficient to study the behavior of a doubly periodic function for values of z in any period-parallelogram, just as in the case of simply periodic functions it is sufficient to examine the function for values of z in any one of the various period-strips. This fact simplifies the discussion of periodic functions. We shall use the term **period-region** to mean either a period-strip or a period-parallelogram according as the given function is simply or doubly periodic.

We shall have no occasion to discuss in this connection functions having more than two independent periods, for it may be shown that a single-valued analytic function can at most be doubly periodic.*

As an illustration of a doubly periodic function, let us consider the function $\wp'(z)$ of Weierstrass, which is defined by the relation (Art. 56)

$$\begin{aligned}\wp'(z) &= -\frac{2}{z^3} - 2\sum{}' \frac{1}{(z-\Omega)^3}\\ &= -2\sum \frac{1}{(z-\Omega)^3}.\end{aligned} \tag{12}$$

As we have seen, this function is holomorphic in the finite region except for the doubly infinite set of Ω-points. Replacing z by $z \pm 2\,\omega_1$, we have from (12)

$$\begin{aligned}\wp'(z \pm 2\,\omega_1) &= -2\sum \frac{1}{(z \pm 2\,\omega_1 - \Omega)^3}\\ &= -2\sum \frac{1}{\{z - (\Omega \mp 2\,\omega_1)\}^3}.\end{aligned}$$

From an examination of Fig. 102, it will be seen that the set of points $z - (\Omega \mp 2\,\omega_1)$ is the same as the set of points $z - \Omega$, the points being taken in another order. The order of the terms in the series defining $\wp'(z)$ is immaterial since the series converges absolutely. Consequently, we have

$$\wp'(z \pm 2\,\omega_1) = \wp'(z).$$

* Forsyth, *Theory of Functions*, 2d Ed., p. 230.

Similarly, it may be shown that

$$\wp'(z \pm 2\,\omega_3) = \wp'(z).$$

Hence, it follows that $\wp'(z)$ remains invariant by the translation

$$\begin{aligned} z' &= z + 2\,m_1\omega_1 + 2\,m_3\omega_3, \\ &= z + \Omega; \end{aligned}$$

that is,

$$\wp'(z + \Omega) = \wp'(z),$$

and the given function is therefore doubly periodic.

THEOREM II. *If a single-valued periodic analytic function $f(z)$ is holomorphic in any given period-region, it is a constant.*

As we have seen, a periodic function takes the same value at the corresponding points of the various period-regions. If the given function is holomorphic in any period-region, then in that region it is continuous and bounded. But if it is bounded in any period-region, it is bounded over the entire complex plane. Hence $f(z)$ has no singular point and is therefore a constant by Theorem XII, Art. 51.

We have also the following theorem.

THEOREM III. *A single-valued periodic analytic function $w = f(z)$ which is not a constant is necessarily a transcendental function of z.*

By Theorem II each period-region must contain at least one singular point. In the case of doubly periodic functions these singular points have the point $z = \infty$ as a limiting point, and consequently that point is an essential singular point. The same result follows in the case of simply-periodic functions if singular points appear in the finite portion of the various period-strips. However, the point at infinity is a point in each period-strip and this point may be the only singular point. In this case the point $z = \infty$ must also be an essential singular point. For, if z_0 is any finite point in one of the period-strips, the function takes then the same finite value $f(z_0)$ at each of the corresponding points $z_0 + k\omega$, where ω is the primitive period of the function and k has the values 1, 2, 3, ... The points have the limiting point $z = \infty$, but the limit of the function as k becomes infinite is the finite value $f(z_0)$. Since we obtain by a particular approach to the point $z = \infty$ a finite limiting value of the function, that point can not be a pole. Since it is a singular point, it must then be an essential singular point. The given function can not be a rational function; for, in that case the point $z = \infty$

can not be an essential singular point. Since all single-valued algebraic functions are rational, and hence can have no other singular points than poles, the given function must be transcendental.

THEOREM IV. *Let $f(z)$ be a single-valued doubly periodic analytic function having only a finite number of singular points in each period-parallelogram. The integral $\int f(z)\,dz$ taken around the contour of a period-parallelogram is zero.*

Denote the contour of a period-parallelogram by C. If z_0 is one of the corner points of the parallelogram, then the other points may be taken as indicated in Fig. 103, z_0 being so chosen that no singular

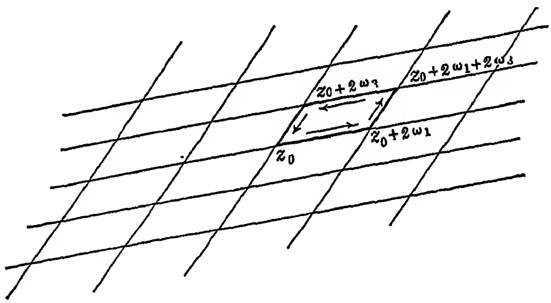

FIG. 103.

points lie on the contour of the period-parallelogram. The integral taken in a positive direction around the contour is

$$\int_C f(z)\,dz = \int_{z_0}^{z_0+2\omega_1} f(z)\,dz + \int_{z_0+2\omega_1}^{z_0+2\omega_1+2\omega_3} f(z)\,dz$$
$$+ \int_{z_0+2\omega_1+2\omega_3}^{z_0+2\omega_3} f(z)\,dz + \int_{z_0+2\omega_3}^{z_0} f(z)\,dz.$$

We can combine the first and third integrals in the right-hand member, and likewise the second and fourth, thus obtaining

$$\int_C f(z)\,dz = \int_{z_0}^{z_0+2\omega_1} \{f(z) - f(z+2\omega_3)\}\,dz - \int_{z_0}^{z_0+2\omega_3} \{f(z) - f(z+2\omega_1)\}\,dz.$$

But as $f(z)$ is doubly periodic having the periods $2\omega_1$, $2\omega_3$, we have

$$f(z+2\omega_3) = f(z),$$
$$f(z+2\omega_1) = f(z),$$

from which it follows that both of the integrals in the second member of the foregoing equation vanish. Hence, we have

$$\int_C f(z)\, dz = 0,$$

as the theorem requires.

By aid of this theorem we can now establish the following theorem.

THEOREM V. *The sum of the residues in a period-parallelogram of a single-valued doubly periodic analytic function having in any period-parallelogram only a finite number of singular points is zero.*

The residue of a function at an isolated singular point was defined as the value of the integral

$$\frac{1}{2\pi i}\int_C f(z)\, dz,$$

where C is a closed path inclosing no other singular point. It has also been shown that this integral taken around the contour of a region containing a finite number of singular points is the sum of the residues of the function at these points. It follows then from Theorem IV that if a function $f(z)$ satisfies the stated conditions, the sum of its residues at the singular points in a period-parallelogram must be zero.

THEOREM VI. *The number of zero-points in any period-parallelogram of a single-valued doubly periodic analytic function, which is not a constant and has no singular points other than poles, is equal to the number of poles in this period-parallelogram, each zero point and each pole being taken a number of times equal to its multiplicity.*

Let $f(z)$ be the given function. Since it is analytic, its zero points are isolated, and hence in any finite region there can only be a finite number. By hypothesis $f(z)$ is periodic. Hence, $f'(z)$ and $\dfrac{f'(z)}{f(z)}$ are likewise both periodic. Moreover, since $f(z)$ has but a finite number of zero points in any finite region, $\dfrac{1}{f(z)}$ can have but a finite number of poles in such a region. Consequently, the quotient $\dfrac{f'(z)}{f(z)}$ has but a finite number of poles in any period-parallelogram. By Theorem V we have

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}\, dz = 0,$$

where γ denotes the contour of any period-parallelogram so chosen that no poles of $\dfrac{f'(z)}{f(z)}$ or of $f(z)$ lie upon γ. But by Theorem X, Art. 54, it follows that the foregoing integral is equal to the number of zero points of $f(z)$ in the period-parallelogram bounded by γ, minus the number of poles of $f(z)$ in the same parallelogram. Since this difference is zero, we have the given theorem.

The foregoing theorem may be generalized as follows:

THEOREM VII. *In any period-parallelogram of a single-valued doubly periodic analytic function $f(z)$, which is not a constant and has in the finite region no other singular points than poles, takes any arbitrary value C a number of times equal to the number of its poles, each pole being taken a number of times equal to its multiplicity.*

To prove this theorem consider the function

$$F(z) = f(z) - C.$$

The conditions of Theorem VI are satisfied by $F(z)$, and hence the number of its zero points is equal to the number of its poles in any period-parallelogram. But the poles of $F(z)$ are at the same time poles of $f(z)$. Moreover, at the zero points of $F(z)$, $f(z)$ takes the value C. Hence, the given function $f(z)$ takes the arbitrary value C in each period-parallelogram a number of times equal to the number of its poles in that parallelogram.

If, as in Theorem VI, our attention is confined to functions having no singular points in the finite part of the plane except poles, it follows that there must be at least one pole in each period-parallelogram. If there is but one pole, its residue must be zero; consequently the order of the pole must be at least two. If only simple poles appear, each period-parallelogram must contain two or more such poles in order that the residue may be zero. Doubly periodic functions having only poles in the finite region form an important class of functions called **elliptic functions.** We shall, however, not consider the special properties of such functions in the present volume.

EXERCISES

1. Show that every rational integral function $f(z)$ is uniquely determined for all complex values of z as soon as $f(x)$ is known for all real values of x from zero to one.

2. Given the function

$$f(z) = \frac{z+3}{z(z^2 - z - 2)}.$$

Represent this function by an infinite series for values of z; (a) within the unit circle about the origin; (b) within the annular region between the concentric circles about the origin having respectively the radii 1 and 2; (c) exterior to the circle of radius 2.

3. Can $\sin z$ ever be greater than one? If so at what points? How do the zeros of $\sin z$ compare with those of $\sin x$, where x is real? What solutions can the equation $\cos z = 1$ have?

4. Given the infinite series

$$1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots + \frac{z^n}{n!} + \cdots.$$

From the form of this series, how do we know that it defines a function $f(z)$ that is holomorphic in the entire finite portion of the complex plane? Knowing the expansion of e^x, how can it be shown that $f(z) = e^z$? How can this same method be used to obtain the expansion of $\sin z$, $\cos z$ from the expansion of $\sin x$, $\cos x$?

5. Given the function

$$f(z) = \frac{e^z}{z^2 + 1}.$$

Is $f(z)$ an analytic function? What is its region of existence? Find a region S in which the function is holomorphic. Locate the singular points and classify the singularities of the function at these points. Does the function have any zero points?

6. Indicate the general form which the infinite expansion of an analytic function $f(z)$ must have in the neighborhood of z_0, if $f(z)$ has (a) a pole of order 3 at z_0, (b) a zero of order 3 at z_0, (c) an essential singularity at z_0.
What is the nature of the function $F(z) = \dfrac{1}{f(z)}$ at z_0 in each case?

7. Locate the zero points of $\sin \dfrac{1}{z}$. What is the nature of this function at the limiting point of these zeros? From this conclusion what can be said of the singular points of $\csc z$? Why is $\csc z$ a transcendental fractional function?

8. Given

$$f(z) = \frac{1}{(z-1)(z-3)}.$$

If this function is expanded in a Taylor series about $z = 2$, how large can the circle of convergence be? Expand the given function in powers of z and determine the circle of convergence.

9. Given the analytic expression

$$\phi(z) = \underset{n=\infty}{L} \frac{1}{1 - z^n}.$$

Show that $\phi(z)$ is an element of two distinct analytic functions, according as we take $|z| < 1$ or $|z| > 1$.

10. Show that a doubly-periodic function $f(z)$ which is an integral transcendental function is a constant.

11. Given the function

$$f(z) = \frac{z^3 + 1}{z^3 - 2z^2 + z - 2}.$$

Determine the zero points and poles. Compute the residue at each of the latter.

12. The function
$$f(z) = \cos^2 z + \sin^2 z$$
has the constant value 1 along the axis of reals. By what property does it follow that it must be equal to one for all finite values of z? What other relations of trigonometric functions can be extended in a similar manner from real to complex values of the variable?

13. Given the integral $\int \frac{z^2}{z^3 + 1}\, dz$ taken along a circle C about the origin as center. How large can the radius of C be taken and have the value of the integral zero? What would the integral be if the radius of C is taken $\frac{1}{2}$ unit larger?

14. Given a function having poles at the points $z = \alpha_1, \alpha_2, \ldots, \alpha_k$, and an essential singularity at $z = \infty$. What kind of a function is it? Write an infinite series which will represent such a function.

15. By use of the theory of residues evaluate the integral
$$\int_{-\infty}^{\infty} \frac{dx}{(x^2 + 2)^2}.$$

16. Give an illustrative example of a single-valued analytic function having (*a*) no singular point, (*b*) no singular point in the finite region and a pole at infinity, (*c*) no singular point other than an essential singular point at infinity, (*d*) a finite number of poles in the finite region and an essential singular point at infinity, (*e*) an infinite number of poles in finite region dense at infinity, (*f*) an isolated essential singular point in the finite region. Classify each function.

17. Show that the infinite product
$$\prod_{k=1}^{\infty} \left(1 + \frac{1}{k^2}\right)$$
converges absolutely.

18. Show without testing the remainder that the expansion of $\tan z$ in terms of z must converge for all values of z less in absolute value than $\frac{\pi}{2}$.

19. If in the deleted neighborhood of $z = z_0$, $f(z)$ is holomorphic and not a constant, and if $f(z)$ takes the same value β at a set of points dense at z_0, then $z = z_0$ is an essential singular point of $f(z)$. Apply this result to testing the nature of $\sin \frac{1}{z}$ at $z = 0$.

HINT: Consider the function $\frac{1}{f(z) - \beta}$.

20. Do the poles of the function
$$f(z) = \frac{3 \sin z + 7 \tan z}{z^2}$$
have a limiting point? What is the nature of $f(z)$ at this point? Show that $f(z)$ can be made to approach an arbitrarily chosen number as z approaches this point. Without computing the integral
$$\int_C f(z)\, dz,$$
where C is the boundary of a region S having $z = \infty$ as an inner point, show that this integral does not vanish.

CHAPTER VIII

PROPERTIES OF MULTIPLE-VALUED FUNCTIONS

59. Fundamental definitions. In the previous chapters we have frequently had occasion to consider single-valued functions whose inverse functions are multiple-valued, that is, are functions having in general two or more values for the same value of the independent variable. The functions

$$w = \sqrt{z}, \qquad w = \log z, \qquad w = \arcsin z$$

are illustrations of multiple-valued functions. In the present chapter we shall consider some of the special properties of such functions. In the consideration of multiple-valued functions we must distinguish between multiple-valued analytic functions and those multiple-valued expressions that represent two or more single-valued analytic functions. Thus

$$w = \sqrt{z^2}$$

is not to be regarded as a multiple-valued analytic function, but rather as two single-valued functions

$$w = z, \qquad w = -z.$$

These two functions are distinct analytic functions rather than elements of the same analytic function, for neither can be deduced from the other by the process of analytic continuation.

In the same way the expression

$$w = \sqrt{1 - \sin^2 z}$$

is not a single multiple-valued analytic function, but represents the two single-valued analytic functions

$$w = \cos z, \qquad w = -\cos z.$$

The expression

$$w = \log e^z$$

represents an infinite number of distinct, single-valued analytic functions, namely

$$w = z, \quad z + 2\pi i, \quad z + 4\pi i, \; \ldots .$$

These functions, like the foregoing functions, are distinct because no one of them can be deduced from another by a process of analytic continuation.

Likewise, the expression $w = \alpha^z$, where $\alpha \equiv a + ib$, is not a multiple-valued analytic function but represents an infinite number of single-valued functions which can not be deduced from each other by the process of analytic continuation. For, we have

$$w = \alpha^z = e^{z \log \alpha},$$

which represents the single-valued analytic functions *

$$e^{z \log \alpha}, \qquad e^{z(\log \alpha + 2\pi i)}, \qquad e^{z(\log \alpha + 4\pi i)}, \; \ldots \; .$$

In mapping by means of multiple-valued functions, we have already seen that the whole of one plane may map upon a portion of the other plane. For example, by means of the function $w = \sqrt{z}$, the whole of the Z-plane may be mapped upon that half of the W-plane for which the real part of $w = u + iv$ is positive. Conversely, by means of the same relation, this half of the W-plane may be mapped upon the whole of the Z-plane. Likewise, the half of the W-plane for which the real part of w is negative may be mapped upon the whole of the Z-plane. If the given function is $w = \sqrt[n]{z}$, the whole of the Z-plane may be mapped upon a sector of the W-plane formed by half-rays drawn from the origin and making angles $\frac{\pi}{n}$, $-\frac{\pi}{n}$, respectively, with the positive axis of reals. In the consideration of multiple-valued functions such as

$$w = \sqrt{z}, \qquad w = \sqrt[n]{z}, \qquad w = \log z,$$

we have thus far restricted the discussion to those values of w which correspond to values of z for which we have

$$-\pi < amp\, z \leqq \pi;$$

that is, we have restricted our consideration to a fundamental region of the inverse function. With this restriction we have been able to consider these functions as though they were single-valued. As we shall see, such an arrangement has the disadvantage that the mapping from at least one of the two planes upon the other is not always continuous.

* For the case where $\alpha = e$, we give to $\log e$ only the value 1 since e^x must appear as a special case of e^z. So also with other logarithms of real numbers.

For example, if we have $w = \sqrt[3]{z}$, then by mapping upon the W-plane the amplitude of z is divided by three; for, putting

$$z = \rho(\cos\theta + i\sin\theta),$$
$$w = \rho'(\cos\theta' + i\sin\theta'),$$

we have from

$$w = \sqrt[3]{z},$$
$$\rho'(\cos\theta' + i\sin\theta') = \rho^{\frac{1}{3}}\left(\cos\frac{\theta}{3} + i\sin\frac{\theta}{3}\right),$$
$$\theta' = \frac{\theta}{3}.$$

All points of the Z-plane, when we consider only the principal amplitude of those points, map into a region I (Fig. 105), bounded by two half-rays from the origin making an angle of 60° and $-60°$, respectively, with the positive half of the axis of reals. The result is as

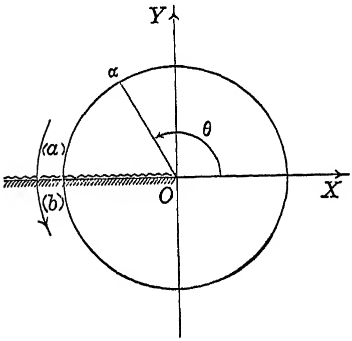

FIG. 104.

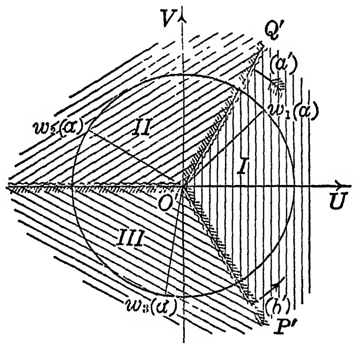

FIG. 105.

though the Z-plane were cut along the negative half of the axis of reals and the whole plane contracted by moving each bank in its own one-half plane through an angle of 120° into the positions of $O'Q'$ and $O'P'$. The mapping from the Z-plane upon the fundamental region of the W-plane is single-valued but not necessarily continuous, as may be seen most clearly by considering a continuous curve in the Z-plane crossing the negative axis of reals. That portion (a) of the curve above the X-axis maps into the curve (a') to the right of $O'Q'$, while the portion (b) below the X-axis maps into (b') to the right of $O'P'$. Consequently, what was a continuous curve becomes after mapping a discontinuous curve as shown in the figure.

Either of the regions II, III might equally well have been selected as the particular region upon which the Z-plane is mapped by placing the proper restrictions upon the variation of the amplitude of z. To any point α in the Z-plane distinct from the origin there correspond then three points in the W-plane when no restriction is placed on the amplitude of z. We may denote these three values by $w_1(\alpha)$, $w_2(\alpha)$, $w_3(\alpha)$.

As z takes all values in the Z-plane, subject to the condition that

$$-\pi < amp\, z \leqq \pi,$$

let w_1 denote the corresponding functional values represented by the points of region I of the W-plane. Likewise, let w_2, w_3 denote the totality of values represented respectively by the points of the regions II, III. Consequently, w_1, w_2, w_3 are single-valued functions of z such that the three taken together give all of the corresponding values of w and z included in the functional relation $w = \sqrt[3]{z}$. We call w_1, w_2, w_3 the three branches of the given function.

Similarly, we may define a branch of any multiple-valued analytic function $w = f(z)$. An assemblage of pair-values (w, z) is said to define a **branch** w_k of such a function if it possesses the following properties:

1. The aggregate of all z-points of the region of existence which enter into consideration must fill a region S_k exactly once.
2. To each point of S_k there corresponds but one value of w.
3. The aggregate of w-points corresponding to points in S_k represents a continuous function of z.

Thus, for the function $w = \sqrt[3]{z}$ each of the regions S_k, $k = 1, 2, 3$, of the Z-plane consists of the whole finite portion of the complex plane with the exception of the point $z = 0$, which is to be thought of as a boundary point of S_k for reasons that will appear later. Taking the negative axis of reals as a portion of the boundary, but nevertheless a part of S_k, it follows that the regions of the W-plane corresponding to the three branches of the function are definitely determined; that is they are the regions I, II, III in Fig. 105. Had the amplitude varied between other limits, say between $-\frac{\pi}{2}$ and $\frac{3\pi}{2}$, the values of w corresponding to values of z in S_k would have been different, and the branches of the given function would have been correspondingly changed.

The branches of a multiple-valued analytic function are single-valued functions of the independent variable, and the totality of the value-pairs (w, z), representing all of the branches taken together, is identical with those of the given function. Whenever the inverse function $z = \phi(w)$ is single-valued, then the domain of the functional values of the given function $w = f(z)$ breaks up into a finite or an infinite number of non-overlapping regions according as $w = f(z)$ has a finite or an infinite number of branches. Moreover, if in this case the given function has no natural boundary, each of these regions of the W-plane may be taken as a fundamental region of the inverse function.

A point is called a **branch-point** of the given analytic function if some of the branches interchange as the independent variable describes a closed path about it. The existence of branch-points is a characteristic of multiple-valued analytic functions as distinguished from multiple-valued expressions, such as $\sqrt{1 - \sin^2 z}$, representing two or more single-valued analytic functions. The point $z = 0$ is a branch-point of the function $w = \sqrt[3]{z}$. For, let z describe a closed path about the origin, beginning at any point α whose amplitude is θ, and let us consider the change that takes place in the value of the function $w = \sqrt[3]{z}$. The initial value of the function is then $w_1(\alpha)$ and is represented by a point in the region I, Fig. 105. After one revolution of z about the origin, the function does not return to its original value, but has changed to a value $w_2(\alpha)$, represented by a point in region II, its amplitude having been increased by $\frac{2\pi}{3}$. After a second revolution of z about the origin the functional value has changed from $w_2(\alpha)$ to a value $w_3(\alpha)$, represented by a point in the region III, and the amplitude of w has changed to $\theta + \frac{4\pi}{3}$. After three revolutions of z about the origin the function again attains its initial value $w_1(\alpha)$.

In the discussion of analytic continuation of single-valued functions, it was shown that when the continuation is taken along any path between two points z_0 and z_1 of a region S in which the function is holomorphic, the same functional value at z_1 is obtained irrespective of the path chosen, provided that path lies wholly within S. For multiple-valued functions a different functional value may be obtained at the terminal point z_1 if the continuation is taken along different paths. This is always the case for some branches of the

function if the two paths are so chosen as to inclose a branch-point; for, the two paths taken together then form a closed path about that point, and from the definition of a branch-point it follows that whenever the independent variable z makes but one circuit of this path some of the branches of the function are interchanged and the initial and final values of the function do not coincide. This property is, by the definition of a branch-point, a characteristic of all paths inclosing such points.

The different branches of a function $w = f(z)$ are connected with one another at the w-points corresponding to the branch-points in the Z-plane. For example, it will be observed that all three branches of the function $w = \sqrt[3]{z}$ come together at the point $w = 0$, which corresponds to the branch-point $z = 0$. As we shall see later, however, not all of the branches of an analytic function need be connected at any particular branch-point. If $k + 1$ branches coincide at a branch-point, that point is then said to be a branch-point of **order** k. In the illustration discussed, the origin is therefore a branch-point of order two. Since the branches of a multiple-valued analytic function are single-valued, such a function may be considered as an aggregate of single-valued functions, so related that their values become identical at the branch-points, each being determined from the others by the process of analytic continuation.

The point at infinity may of course be a branch-point. To examine its nature we may make use of the usual substitution $z = \frac{1}{z'}$, and examine the nature of the transformed function at the origin.

In the neighborhood of a branch-point, the mapping by means of an analytic function ceases to be conformal; for, from the foregoing discussion it follows that angles are not preserved at such a point. In the case of $w = \sqrt[3]{z}$, for example, it was seen that at the branch-point $z = 0$ angles are divided by three in mapping from the Z-plane upon the W-plane. Suppose we have given a multiple-valued function $w = f(z)$, whose inverse function $z = \phi(w)$ is single-valued. If this inverse function fails to map the W-plane in the neighborhood of w_0, where $w_0 = f(z_0)$, conformally upon the Z-plane, then it follows that the derived function $\phi'(w)$ must either be zero or become infinite for $w = w_0$. This fact enables us to formulate a convenient test for finding the branch-points of a function whose inverse function is single-valued. Such a criterion is given in the following theorem, where w_0 is a finite point.

THEOREM. *Given a multiple-valued analytic function $w = f(z)$, whose inverse function $z = \phi(w)$ is single-valued. If the derived function $\phi'(w)$ has a zero point of order k, or a pole of order $k + 2$, at w_0, then $f(z)$ has a branch-point of order k at the corresponding point z_0.*

Let us suppose that $\phi'(w)$ has at w_0 a zero point of order k. We may then write

$$\phi'(w) = (w - w_0)^k F_1(w), \tag{1}$$

where k is a positive integer and $F_1(w)$ is holomorphic in the neighborhood of w_0 and different from zero for $w = w_0$. It follows that the factor $(w - w_0)$ enters into $\phi(w) - \phi(w_0)$ to a degree one higher, that is to the degree $k + 1$, thus giving

$$\phi(w) - \phi(w_0) = (w - w_0)^{k+1} F_2(w), \tag{2}$$

where $F_2(w)$ is also holomorphic in the neighborhood of w_0 and different from zero for $w = w_0$. Let A be the principal $(k + 1)^{st}$ root of $F_2(w_0)$ and let $\chi(w)$ be a function having the point w_0 as a regular point such that $\chi(w_0)$ is equal to A. The function $\chi(w)$ can be so determined that for values of w in the neighborhood of w_0 we have

$$\{\chi(w)\}^{k+1} = F_2(w).$$

From (2) we now have

$$\phi(w) - \phi(w_0) = \{(w - w_0)\,\chi(w)\}^{k+1}.$$

As a matter of convenience, we introduce the auxiliary function

$$\tau = (w - w_0)\,\chi(w). \tag{3}$$

Let us now suppose the τ-plane to be mapped upon the W-plane by means of this relation. The point $\tau = 0$ corresponds to the point $w = w_0$, and the mapping is conformal in the neighborhood of $\tau = 0$; because we have

$$\left.\frac{dw}{d\tau}\right]_{\tau=0} = \left.\frac{1}{\dfrac{d\tau}{dw}}\right]_{w=w_0} = \left.\frac{1}{\chi(w) + (w - w_0)\,\chi'(w)}\right]_{w=w_0} = \frac{1}{\chi(w_0)},$$

where $\dfrac{1}{\chi(w_0)}$ is finite and different from zero, since $\chi(w_0)$ is the principal $(k + 1)^{st}$ root of $F_2(w_0)$, which is different from zero.

If we now map the finite portion of the Z-plane upon the τ-plane by means of the relation

$$\tau = \sqrt[k+1]{\phi(w) - \phi(w_0)} = \sqrt[k+1]{z - z_0}, \tag{4}$$

the neighborhood of z_0 maps into the neighborhood of the origin in the τ-plane. The two substitutions (3) and (4) are together equivalent to mapping at once from the Z-plane to the W-plane by means of the relation

$$z = \phi(w),$$

or

$$w = f(z).$$

From what has been said about mapping by means of the relation $w = \sqrt[n]{z}$, it follows at once from (4) that $z = z_0$ is a branch-point of the order k, which shows that the condition stated in the theorem is valid if w_0 is a zero point of order k of $\phi'(w)$.

Let us now consider the case where w_0 is a pole of order $k + 2$ of $\phi'(w)$. The corresponding z-point is the point at infinity. For this case equation (1) takes the form

$$\phi'(w) = \frac{F_1(w)}{(w - w_0)^{k+2}}.$$

Applying the same method as employed in the foregoing discussion, it follows at once that $f(z)$ has a branch-point of order k at $z = \infty$. The details of the proof are left as an exercise for the student.

The difficulties in representing multiple-valued functions in the foregoing manner arise for the most part from the fact that a continuous curve in the one plane does not always correspond to a continuous curve in the other. Such difficulties can be easily avoided by means of a simple device known as a Riemann surface, consisting of n sheets connected with one another in a definite manner depending upon the character of the function. The nature of such a surface can perhaps be most readily made clear by means of an illustrative example.

Let us consider the Riemann surface for the function

$$w = \sqrt[3]{z}.$$

It has already been pointed out (Fig. 105) that the whole of the Z-plane maps by means of this function into any one of the three regions I, II, III of the W-plane. To each z-point correspond in general three w-points, one in each of the regions I, II, III. Suppose we think of the Z-plane as consisting of three sheets (Fig. 106) connected with one another at $z = 0$ and along the negative axis of reals. As z describes a circuit about the branch-point $z = 0$, suppose it passes from one sheet to another upon crossing the negative axis of reals and that the variable point enters the various sheets

in the same order as the branches of the function were permuted in the previous discussion when z described a closed path about the same point. Let the points of the region I be placed into correspondence with those of the first sheet of the Z-surface, the points of region II with those of the second sheet, and the points of region III with those of the third sheet. Corresponding to the three points $w_1(\alpha)$, $w_2(\alpha)$, $w_3(\alpha)$, of the W-plane we have then a point α in each of the three sheets of the Z-plane, denoted by $\alpha^{(1)}$, $\alpha^{(2)}$, $\alpha^{(3)}$, respectively. The branch-point $z = 0$ is not to be regarded as a point of the region of existence of the given function, but is to be considered as belonging to its boundary. The same is to be said of any branch-point so far as the sheets of the surface affected are concerned. Each sheet of the Riemann surface, like the ordinary complex plane is to be regarded as closed at infinity. By aid of the Riemann surface there is thus established a one-to-one correspondence between the points of the W-plane and the three-sheeted Z-plane. Let z start from a point $\alpha^{(1)}$ in the first sheet and describe a continuous path about the branch-point $z = 0$. By going once around the origin, z does not return to $\alpha^{(1)}$ but changes to the point $\alpha^{(2)}$ in the second sheet. As z crosses the negative axis of reals in passing from $\alpha^{(1)}$ to $\alpha^{(2)}$, the corresponding w-point crosses the line $O'Q'$ and passes from the point $w_1(\alpha)$ in the region I to the point $w_2(\alpha)$ in the region II. By a second revolution about $z = 0$, z again crosses the negative axis of reals and passes into the third sheet reaching the point $\alpha^{(3)}$. At the same time the corresponding w-point passes from the region II to the region III reaching the position $w_3(\alpha)$. After the third revolution about the branch-point $z = 0$, z returns to the first sheet upon crossing the negative axis of reals and again takes the value $\alpha^{(1)}$. At the same time w passes across the line $O'P'$ and is again in region I, finally assuming its initial value $w_1(\alpha)$ as z coincides with $\alpha^{(1)}$. Had the revolution taken place in the opposite direction about the branch-point, the order in

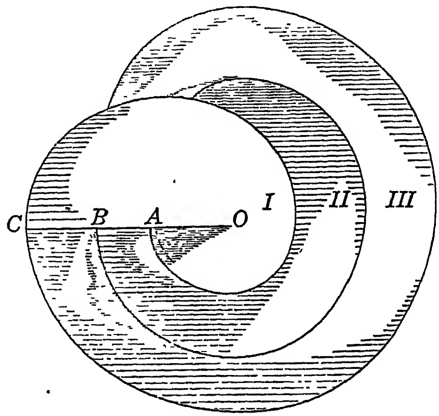

Fig. 106.

which z would have changed sheets would have been from the first to the third, from the third to the second, and finally from the second to the first. The negative axis of reals is called a **branch-cut,** for as z passes over it in either direction the functional values change from one branch to another. A cross-section of the Z-plane shows the connection of the three sheets. If taken perpendicular to the negative axis of reals, the cross-section showing the intersection of sheets along that portion of the axis of reals, as viewed from the origin, appears as shown in Fig. 107.

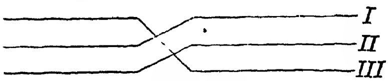

FIG. 107.

As will be seen, the advantage of introducing a Riemann surface in place of the single-sheeted complex plane is that every continuous curve on the Z-plane maps by a multiple-valued analytic function into a continuous curve on the W-plane, and conversely. This relation between the two planes enables us to bring to the consideration of multiple-valued analytic functions all such processes as integration, analytic continuation, etc., depending on a continuous path being drawn from one point to another. In case w and z are each a multiple-valued function of the other, then both planes are replaced by a Riemann surface whose character is determined by the nature of the function under consideration. It is often convenient in such cases to introduce an auxiliary plane, whereby the one Riemann surface may be mapped upon this single-sheeted complex plane and this plane in turn mapped upon the second Riemann surface. In the following articles we shall consider more in detail the various properties of Riemann surfaces.

60. Riemann surface for $w^3 - 3w - 2z = 0$. For any value of z there are in general three values of w; evidently, therefore, there are three branches of the given function, which we shall denote by w_1, w_2, w_3. The branch-points can be at once determined by aid of the theorem of Art. 59. For z is a single-valued function of w, and moreover we have

$$\frac{dz}{dw} = \frac{3}{2}(w^2 - 1),$$

which has a zero point of order one at $w = +1, -1$, and a pole of order two at $w = \infty$. Consequently, the given function must have simple branch-points at $z = -1, +1$, and a branch-point of order two at $z = \infty$, these three values corresponding, respectively, to $w = +1, -1, \infty$. That there can be no other branch-points than

these follows from the fact that there are no other values of z for which two or more of the branches become equal. The Z-plane then must consist of a three-sheeted Riemann surface, the three sheets all being connected at $z = \infty$, and two of them at $z = 1$, and likewise two at $z = -1$.

The manner in which the sheets of the Riemann surface are connected at the three branch-points and a convenient way for drawing the necessary branch-cuts can be determined by examining the manner in which the Z-plane may be mapped upon the regions of the W-plane corresponding to the three branches of the function. The branches w_1, w_2, w_3 can be expressed in terms of z by solving the given equation

$$w^3 - 3w - 2z = 0 \tag{1}$$

for w by means of Cardan's solution of the cubic.

The general equation of the third degree can be reduced to the form

$$w^3 + 3Hw + G = 0,$$

and Cardan's solution applies equally well whether H, G, are real or complex.* The three roots of this equation are then

$$w_1 = p + q, \qquad w_2 = \omega p + \omega^2 q, \qquad w_3 = \omega^2 p + \omega q, \tag{2}$$

where ω is one of the imaginary cube roots of unity and

$$p = \sqrt[3]{-\frac{G}{2} + \sqrt{\frac{G^2}{4} + H^3}}, \qquad q = \sqrt[3]{-\frac{G}{2} - \sqrt{\frac{G^2}{4} + H^3}},$$

subject, however, to the condition

$$pq = -H.$$

For the case under consideration, we have $H = -1$, $G = -2z$, and hence

$$p = \sqrt[3]{z + \sqrt{z^2 - 1}}, \qquad q = \sqrt[3]{z - \sqrt{z^2 - 1}}, \tag{3}$$

subject to the condition that $pq = 1$.

We shall now introduce the auxiliary relation

$$z = \cos 3\tau, \tag{4}$$

by means of which we obtain from (3)

$$p = \sqrt[3]{\cos 3\tau + \sqrt{\cos^2 3\tau - 1}} = \sqrt[3]{\cos 3\tau + i\sqrt{\sin^2 3\tau}},$$

$$q = \sqrt[3]{\cos 3\tau - \sqrt{\cos^2 3\tau - 1}} = \sqrt[3]{\cos 3\tau - i\sqrt{\sin^2 3\tau}}.$$

* See Serret, *Cours d'algèbre supérieure*, 3rd Ed., Vol. II, p. 427.

The radical $\sqrt{\sin^2 3\tau}$ must be taken with the same sign, say the plus sign, in both p and q. Both p and q have three values, since each is the cube root of a given number. Any of these values may be chosen which satisfy the added condition $pq = 1$. We may, therefore, put

$$p = \sqrt[3]{\cos 3\tau + i \sin 3\tau} = \cos\tau + i \sin\tau = e^{i\tau},$$
$$q = \sqrt[3]{\cos 3\tau - i \sin 3\tau} = \cos\tau - i \sin\tau = e^{-i\tau}.$$

Remembering that ω is an imaginary cube root of unity, we may write

$$\omega = e^{\frac{2\pi i}{3}}, \qquad \omega^2 = \frac{1}{\omega} = e^{-\frac{2\pi i}{3}}.$$

The three branches w_1, w_2, w_3 of the given function may now be expressed in terms of τ as follows:

$$w_1 = e^{i\tau} + e^{-i\tau} = 2\cos\tau, \tag{5}$$

$$w_2 = e^{i\left(\tau+\frac{2\pi}{3}\right)} + e^{-i\left(\tau+\frac{2\pi}{3}\right)} = 2\cos\left(\tau + \frac{2\pi}{3}\right), \tag{6}$$

$$w_3 = e^{i\left(\tau-\frac{2\pi}{3}\right)} + e^{-i\left(\tau-\frac{2\pi}{3}\right)} = 2\cos\left(\tau - \frac{2\pi}{3}\right). \tag{7}$$

The mapping from the Z-plane upon the W-plane by means of the given relation now reduces to mapping the Z-plane upon the τ-plane by means of the inverse of the function given in (4) and then mapping the τ-plane upon the W-plane by means of the three relations (5), (6), (7). As we shall see, the three branches map into distinct portions of the W-plane, which come together, however, at the points corresponding to the branch-points $z = -1, +1, \infty$. We have a choice of the fundamental region in the τ-plane, and it serves our purpose to take that region bounded by the axis of imaginaries and the line parallel to it and cutting the axis of reals at the point $\left(\frac{\pi}{3}, 0\right)$. The whole Z-plane may be mapped exactly once upon this fundamental region of the τ-plane. This fundamental region for τ in turn may be mapped by means of the relations (5), (6), (7) into each of three definite regions I, II, III of the W-plane.

The results of the mapping from the Z-plane upon the W-plane are exhibited in Figs. 108 and 109. The details are left as an exercise for the student. By our choice of the fundamental region in the τ-plane, the Z-plane is mapped by means of the branch w_1 into

the region I, the upper half of the Z-plane mapping into the portion of this region above the axis of reals, and the lower half into the portion below that axis. In a similar manner, the whole of the

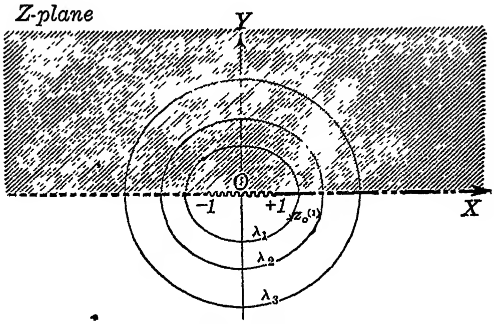

FIG. 108.

Z-plane maps into the region II by means of w_2 and into III by means of w_3 as indicated. Corresponding to the branch-point $z = 1$, we have $w = -1$ and at this point the branches w_2, w_3, become

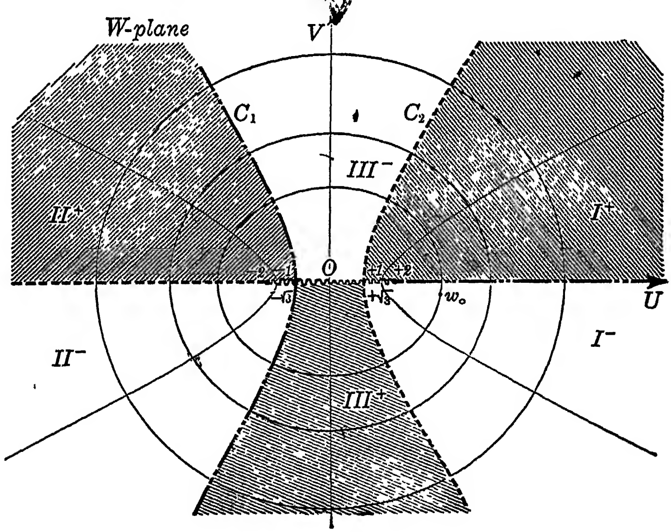

FIG. 109.

identical. For $z = -1$, we have $w = 1$, and at this point, as may be seen from Fig. 109, the branches w_1, w_3, become identical.

Let us now think of the Z-plane as consisting of three sheets. To the first sheet we associate the values of w in I, and to the second

and third sheets we then associate the values of w in II and III, respectively. As z traverses a small closed circuit about $z = -1$ in the positive direction starting from a point in the first sheet, w will pass from I to III and consequently z passes from the first sheet to the third sheet of the three-sheeted Riemann surface constituting the Z-plane. By going about $z = -1$, w never passes into II, since only III and I come together at the corresponding point $w = 1$. In a similar manner, it will be seen that as z traverses once a small closed circuit about $z = 1$, beginning at a point in the second sheet, w passes from II into III and upon continuing a second time about this branch-point w returns to its original position in II. In the neighborhood of this branch-point it is impossible for z to pass from the second or third sheet into the first sheet, since the region I is not associated with II, III at the corresponding point $w = -1$. From Fig. 109, it is apparent that all three branches w_1, w_2, w_3, become identical at the branch-point $z = \infty$. The same result can be obtained analytically by putting $z = \dfrac{1}{z'}$ and examining the transformed function for $z' = 0$. The point $z = \infty$ is therefore to be considered as belonging to the boundary of the region of existence on the Riemann surface and not as an inner point of that region.

It is now convenient to take as the branch-cuts that portion of the axis of reals, Fig. 108, extended from $z = 1$ indefinitely toward the right and from $z = -1$ indefinitely toward the left. The first of these segments maps into the boundary curve C_1 and the second into C_2 passing through $w = -1$, $w = 1$, respectively, Fig. 109. Along the axis of reals between $z = -1$ and $z = 1$, there is then no

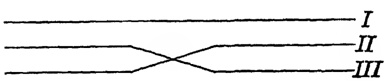

FIG. 110.

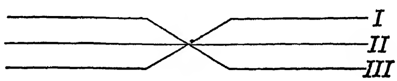

FIG. 111.

connection between the various sheets of the Z-plane, as will be seen by observing the connection between the branches of the function along that portion of the U-axis into which this portion of the X-axis maps. To the right of the point $z = 1$, the various sheets of the Riemann surface are connected as shown in Fig. 110.

To the left of the point $z = -1$, the sheets are connected as shown in Fig. 111.

The discussion of the Riemann surface which replaces the Z-plane

for the given function is now complete. Any continuous curve upon this surface maps into a continuous curve upon the W-plane. For example, the closed curves upon the Riemann surface, of which the ellipses λ_1, λ_2, λ_3 about the points $z = 1, -1$ as foci are the traces, map into ellipses in the W-plane. If the variable z describes the ellipse λ_1, commencing with a point $z_0^{(1)}$ in the first sheet, then w describes a corresponding path beginning at w_0 lying in I^-. As z crosses the positive X-axis the point continues in the first sheet and w passes into I^+. Upon crossing the negative X-axis, the point z passes from the first sheet into the third sheet and w passes from I^+ into III^-. When z has completed one revolution, it is still in the third sheet and we denote its position by $z_0^{(3)}$. By a second revolution of z about λ_1, z passes from the third sheet to the second upon passing across the positive X-axis and remains in the second sheet as it crosses the negative X-axis ending with the value $z_0^{(2)}$. By a third revolution about λ_1 the point passes from the second sheet to the third sheet upon crossing the positive X-axis and again from the third to the first sheet upon passing the negative X-axis, ending with the original position $z_0^{(1)}$.

61. Riemann surface for $w = \sqrt{z - z_0} + \sqrt[3]{\dfrac{z^2}{z - z_1}}$. When rationalized, the given function is seen to be an algebraic function of the sixth degree in w. For each value of z there are then in general six distinct values of w, which we shall denote by $w_1, w_2, w_3, w_4, w_5, w_6$. When the branch-cuts have been drawn upon the Riemann surface, the aggregate of w-values given by $w_1, w_2, w_3, w_4, w_5, w_6$ in terms of z become definite and are respectively the six branches of the function. Considered as functions of z, we may, therefore, refer to them as the branches of the given functions. First of all we shall attempt to discover the branch-points. These points are to be found among those values of z, finite or infinite, for which two or more of the values of w become identical. Not all such points need be branch-points, but no other points can be. We shall accordingly examine the points $z = 0, z_0, z_1, \infty$. We can not make use of the theorem of Art. 59, since z is not a single-valued function of w. We can, however, determine which of these points are branch-points by allowing z to describe an arbitrarily small circuit about each and observing whether the function returns to its initial value.

For convenience we put

$$\sqrt{z - z_0} = \tau_1, \qquad \sqrt[3]{z^2} = \tau_2, \qquad \sqrt[3]{z - z_1} = \tau_3.$$

The six values of w may be written in the form

$$\left.\begin{array}{ll} w_1 = \tau_1 + \dfrac{\tau_2}{\tau_3}, & w_4 = -\tau_1 + \dfrac{\tau_2}{\tau_3}, \\ w_2 = \tau_1 + \omega\dfrac{\tau_2}{\tau_3}, & w_5 = -\tau_1 + \omega\dfrac{\tau_2}{\tau_3}, \\ w_3 = \tau_1 + \omega^2\dfrac{\tau_2}{\tau_3}, & w_6 = -\tau_1 + \omega^2\dfrac{\tau_2}{\tau_3}, \end{array}\right\} \tag{1}$$

where ω is an imaginary cube root of unity. Let z describe a circuit about $z = 0$, taken sufficiently small to exclude both z_0 and z_1. Since $z^{\frac{2}{3}} = \tau_2$, it follows that by one revolution of the circuit by z, τ_2 rotates through an angle of $\dfrac{4\pi}{3}$, which is equivalent to multiplying τ_2 by ω^2. Hence, since τ_1 and τ_3 return to their original values after each revolution, w_1 is changed into w_3, w_3 into w_2, w_2 into w_1, w_4 into w_6, w_6 into w_5, w_5 into w_4. By a second revolution of the circuit a similar change takes place, and upon a completion of the third revolution, the original values are restored. The results may be exhibited as follows:

Before z changes, we have	$w_1, w_2, w_3;\ w_4, w_5, w_6;$
after one revolution, we have	$w_3, w_1, w_2;\ w_6, w_4, w_5;$
after two revolutions, we have	$w_2, w_3, w_1;\ w_5, w_6, w_4;$
after three revolutions, we have	$w_1, w_2, w_3;\ w_4, w_5, w_6.$

The point $z = 0$ is therefore a branch-point. It will be seen that the six branches form two sets, w_1, w_2, w_3 and w_4, w_5, w_6, each of which is cyclically permuted by successive revolutions of z about $z = 0$. By successive revolutions about this point, the branches constituting the first set do not pass into those of the second.

In a similar manner we can test the point $z = z_1$. By each revolution of z about this point τ_1, τ_2 remain unchanged, while τ_3 is multiplied by ω. Remembering that the factor τ_3 appears in the denominator and that $\dfrac{1}{\omega} = \omega^2$, $\dfrac{1}{\omega^2} = \omega$, $\dfrac{1}{\omega^3} = 1$, we have, as the result of the successive revolutions about z_1, the same cyclical interchange of the various values of w as about the origin. Consequently, the point $z = z_1$ is likewise a branch-point.

As z describes an arbitrarily small circuit about z_0, the values of τ_2, τ_3 remain unchanged but τ_1 changes sign by each revolution. The results are as follows:

Before z changes, we have	$w_1, w_2, w_3, w_4, w_5, w_6;$
after one revolution, we have	$w_4, w_5, w_6, w_1, w_2, w_3;$
after two revolutions, we have	$w_1, w_2, w_3, w_4, w_5, w_6.$

Consequently, $w = z_0$ is also a branch-point. All of the branches are affected, but they form three sets, each set including two branches, namely:

$$w_1 \text{ and } w_4, \qquad w_2 \text{ and } w_5, \qquad w_3 \text{ and } w_6.$$

To examine the point $z = \infty$, we put $z = \dfrac{1}{z'}$ and have

$$w' = \sqrt{\frac{1}{z'} - z_0} + \sqrt[3]{\frac{\dfrac{1}{z'^2}}{\dfrac{1}{z'} - z_1}}$$

$$= \frac{\sqrt{1 - z_0 z'}}{\sqrt{z'}} + \frac{1}{\sqrt[3]{1 - z_1 z'}\,\sqrt[3]{z'}}.$$

Putting

$$\sqrt{1 - z'z_0} = \lambda_1, \qquad \sqrt{z'} = \lambda_2, \qquad \sqrt[3]{1 - z_1 z'} = \lambda_3, \qquad \sqrt[3]{z'} = \lambda_4,$$

we have for the six values of w'

$$w_1' = \frac{\lambda_1}{\lambda_2} + \frac{1}{\lambda_3\lambda_4}, \qquad w_4' = -\frac{\lambda_1}{\lambda_2} + \frac{1}{\lambda_3\lambda_4},$$

$$w_2' = \frac{\lambda_1}{\lambda_2} + \omega\frac{1}{\lambda_3\lambda_4}, \qquad w_5' = -\frac{\lambda_1}{\lambda_2} + \omega\frac{1}{\lambda_3\lambda_4},$$

$$w_3' = \frac{\lambda_1}{\lambda_2} + \omega^2\frac{1}{\lambda_3\lambda_4}, \qquad w_6' = -\frac{\lambda_1}{\lambda_2} + \omega^2\frac{1}{\lambda_3\lambda_4}.$$

As z' describes an arbitrarily small circuit about the origin, λ_1, λ_3 are not changed, λ_2 changes sign and λ_4 is multiplied by the factor ω, where ω, as before, is an imaginary cube root of unity. The results of successive revolutions about this circuit are shown in the following table

Before z' changes, we have	$w_1', w_2', w_3', w_4', w_5', w_6'$;
after one revolution,	$w_6', w_4', w_5', w_3', w_1', w_2'$;
after two revolutions,	$w_2', w_3', w_1', w_5', w_6', w_4'$;
after three revolutions,	$w_4', w_5', w_6', w_1', w_2', w_3'$;
after four revolutions,	$w_3', w_1', w_2', w_6', w_4', w_5'$;
after five revolutions,	$w_5', w_6', w_4', w_2', w_3', w_1'$;
after six revolutions,	$w_1', w_2', w_3', w_4', w_5', w_6'$.

It follows that $z' = 0$ and therefore $z = \infty$ is a branch-point where all of the sheets of the Riemann surface are associated with one another.

The Riemann surface constituting the Z-plane has then the branch-points $z = 0, z_0, z_1, \infty$. No two of the finite branch-points can be

connected with a branch-cut. At first thought, it may seem possible to connect the two finite branch-points $z = 0$ and $z = z_1$ with a branch-cut since the same sheets are connected in the same way at these two points. This, however, is not possible; for, by successive revolutions about each of these points the sheets would be connected along such a branch-cut as shown in Fig. 112.

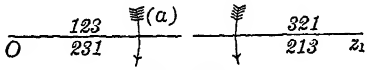

FIG. 112.

Then, by crossing the branch-cut, to the right of $z = 0$ in the direction of arrow (a) we would pass, for example, from the first sheet to the second sheet, while in passing across the branch-cut to the left of $z = z_1$ in the same direction we would pass from the first sheet to the third sheet. Such connection of the sheets along a branch-cut is impossible since the connection must be between the same sheets along its entire length. It is at once apparent that in order to connect any two finite branch-points with a branch-cut, the cyclical interchange of branches must involve the same sheets but in reverse order. We can, however, always draw branch-cuts from any branch-point to the point $z = \infty$; that is, the branch-cuts may be taken as lines extending out indefinitely from the branch-points. Drawing these lines in any convenient manner, the six branches of the given function are fully determined by the six definite aggregates of value-pairs (w, z), such that to the values of z associated with each sheet of the Riemann surface, there is a definite branch of the function. In this case, however, there are no corresponding fundamental regions in the W-plane, since the inverse function is not single-valued. The general discussion of the Riemann surface required for the Z-plane is now complete.

62. Riemann surface for $w = \log z$. The logarithm is a function having an infinite number of branches. As we have seen the logarithm of z may be written in the form

$$w = \log z = \log \rho + i\theta, \qquad (1)$$

where

$$z = \rho(\cos \theta + i \sin \theta).$$

By means of the relation (1) the whole of the Z-plane maps into any one of the strips

$$(2k - 1)\pi < v \leqq (2k + 1)\pi$$

of the W-plane parallel to the axis of reals. Conversely, any one of this infinite number of strips maps into the whole of the Z-plane.

We may now replace the Z-plane by a Riemann surface having an infinite number of sheets. The point $z = 0$ is a branch-point of the surface; for, as we see from (1), every revolution of z in a positive direction about a circle having the origin as a center leaves ρ unchanged but increases θ by 2π, thus changing the value of w. This change continues indefinitely by successive revolutions. We may take the negative half of the axis of reals as the branch-cut, so that every time the variable point crosses this part of the axis it passes from one sheet to the next succeeding or next preceding sheet, according as θ is increasing or decreasing. A cross-section of the surface perpendicular to the negative half of the axis of reals, as seen from the origin, is then of the form shown in Fig. 113. In the same way the point $z = \infty$ may be shown to be also a branch-point of infinitely high order.

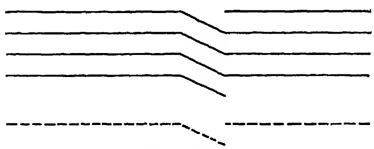

FIG. 113.

The inverse function of $w = \log z$ is, as we know, $z = e^w$. The trigonometric functions were defined in terms of the exponential function, so that the general character of the Riemann surfaces connected with the inverse trigonometric functions may be easily deduced by aid of the logarithmic function.* In the discussions to follow we shall have occasion to consider for the most part only Riemann surfaces having a finite number of sheets.

63. Branch-points, branch-cuts. Having discussed some typical illustrations of Riemann surfaces, we shall now consider some of the more important properties of branch-points and branch-cuts and their relation to the Riemann surfaces needed in the representation of multiple-valued functions.

The branch-points of a function are always to be found among those points corresponding to the values of the independent variable for which two or more of the values of the function become equal. This common value of the various branches of the function may be finite or infinite. As we have seen, not all of the various sheets of a Riemann surface need be connected at any particular point; for, they may be associated in distinct sets at a branch-point as was the case in the points $z = z_0$, $z = 0$, $z = z_1$ for the function discussed in Art. 61. It is essential, however, that all of the sheets be so con-

* Cf. Fouët, *Leçons élémentaires sur la théorie des fonctions analytiques*, 2d Ed., Tome II, p. 128 et seq.

nected at the various branch-points that the entire surface forms a connected whole and it is possible to proceed along a continuous path from any point to any other point upon the surface. As already pointed out, the branch-points themselves, while points of the Riemann surface, are not to be regarded as points of the region of existence of the given function, but as boundary points of that region. The region of existence is then, as with single-valued analytic functions, to be considered an open region, which in the case of multiple-valued functions, however, extends to the several sheets composing the Riemann surface. It is convenient to speak of the branch-points as points of the function in the sense that their character aids in describing the character of the function.

Whenever the inverse function is single-valued, we have in the theorem of Art. 59 a convenient test for locating the branch-points and determining their order. This test is not, however, valid when we have as in Art. 61 a function whose inverse is multiple-valued. In that case it is necessary to test each point where some of the values of the function become identical by permitting the independent variable to traverse an arbitrarily small closed circuit about the point and observe whether the corresponding values of the function return to the initial value.

In the illustrative examples discussed, it was observed that the branches of a function associated with each other at a branch-point were cyclically interchanged by successive revolutions about the point. As a matter of fact, it is always true that all of the branches affected at a branch-point can be so arranged that by successive circuits of the independent variable about the point these branches are cyclically interchanged. The cycle may include all or only a portion of the branches affected. In the latter case there may be two or more sets each of which undergoes a cyclical interchange as the independent variable traverses a circuit about the branch-point a suitable number of times. To show that the branches of the function are interchanged as stated, suppose we let the branches affected be $w_1, w_2, w_3, \ldots$. Then as z describes a small circuit about the branch-point, w_1 must change into some other branch, say w_2. By a second revolution about the point, w_2 can not remain unchanged; for, in that case, tracing the circuit in a reverse order would not restore the initial value w_1, as it should do. It is possible that this second revolution changes the branch w_2 back into w_1, in which case w_1, w_2 constitute a complete cycle. If this is not the case then w_2

must change into some one of the remaining branches, say w_3. As before, when z describes the circuit about the branch-point a third time w_3 can not remain unchanged and can either change back into w_1 or into some one of the remaining branches, say w_4. In the first case, the three branches w_1, w_2, w_3 constitute by themselves a set which cyclically interchange. In the other case, w_4 changes in a similar way into w_5, say, by another revolution about the branch-point. This method of interchange may involve all of the branches in one set, or in several such sets. There may be other sheets that are not affected at this particular branch-point, the connection with one or more of the remaining sheets being made at another branch-point.

The same sheets may be affected at two branch-points. If the order of the cyclic permutation of the sheets in the one case is the reverse of what it is in the other, then it is always possible to connect the two points with a branch-cut; for, by crossing this cut at any point, the variable passes from one sheet into another particular sheet of the cycle. Instead of connecting the branch-points with each other it is always possible to connect the various branch-points with the point at infinity with branch-cuts. Thus far in the discussion the various branch-cuts have been taken as straight lines. This restriction is unnecessary, however, as the choice of a particular curve for the branch-cut is a purely arbitrary convention. A branch-cut is, however, always to be taken so that it has no double-points, that is, the cut must not intersect itself. While the particular aggregate of value-pairs (w, z) constituting the various branches are changed by varying the branch-cut, the connection of the branches with each other at the branch-points remains unchanged. The following illustration will make this statement clear.

Ex. 1. Discuss the branch-cuts of the Riemann surface required for the function $w = \sqrt{\dfrac{i-z}{i+z}}$.

The Z-plane is a double-sheeted Riemann surface and the points $z = i$ and $z = -i$ are simple branch-points. The corresponding functional values are respectively $w = 0$ and $w = \infty$. The result of mapping the Z-plane upon the W-plane is exhibited in Figs. 114, 115. Any curve in the Z-plane joining the points i and $-i$ may be taken as a branch-cut and will map into a curve in the W-plane joining the points $w = 0$ and $w = \infty$. For example, if we select that portion of the axis of imaginaries joining the two branch-points, then the corresponding curve in the W-plane is the positive half of the axis of reals. To every point in the Z-plane, however, correspond two points in the W-plane and

this same portion of the axis of imaginaries also maps into the negative half of the axis of reals in the W-plane. Consequently, in this case the two branches w_1, w_2 of the function corresponding to the values of z in the first and second sheets of the Z-plane are represented by points in the upper and the lower half of the W-plane respectively. If we take any circle through i and $-i$ and select that portion (2) lying to the right of the line $x = 0$ as the branch-cut, then

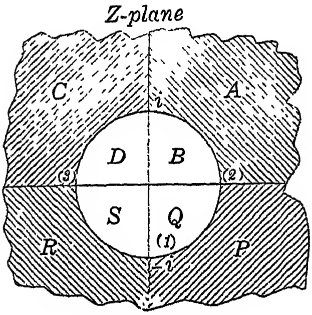

FIG. 114.

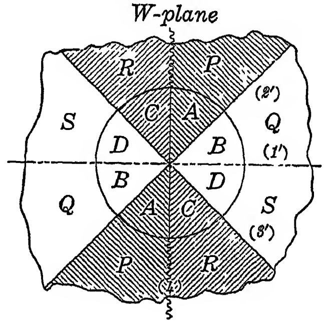

FIG. 115.

the corresponding curve in the W-plane is a half-ray (2′) proceeding from the origin. The same curve (2) also maps into the half-ray making an angle of 180° with (2′). The two half-rays taken together divide the W-plane into two regions and the points in these two regions represent the values of w corresponding to the two branches of the given function. If instead of a circle through i and $-i$ any ordinary curve had been taken as a branch-cut, it would map into a curve joining the points $w = 0$ and $w = \infty$ and again into a congruent curve, which may be obtained by rotating the first curve through an angle of 180°. These two curves taken together constitute a continuous curve dividing the W-plane into two regions whose points give the values of the two branches w_1, w_2 of the function corresponding to the two sheets of the Riemann surface constituting the Z-plane. Again we may select as the branch-cuts of the Riemann surface any curves joining the two branch-points i and $-i$ with the point $z = \infty$, for example those portions of the axis of imaginaries exterior to the circle of unit radius about the origin. This selection likewise leads to a division of the W-plane into two regions and corresponding branches of the function.

From this discussion, it will be seen that the branch-cuts can be selected in a variety of ways and may or may not be straight lines. By the selection of the branch-cuts particular values of the function constituting the various branches are determined. The number and the association of such branches are determined by the character of the function itself. While the selection of the branch-cut is arbitrary, there is often an advantage in selecting it in a particular manner. For example, in the discussion of the Riemann surface for $w = \sqrt[3]{z}$,

we chose the negative half of the axis of reals as the branch-cut in order that one branch of the function should correspond to the principal value of the amplitude of z. Again in the logarithmic function the negative half axis was chosen for the same reason. In discussing the inverse of a periodic function, it is likewise an advantage to select the branch-cuts so that the previously determined fundamental regions shall correspond to single sheets of the Riemann surface rather than conversely.

The following theorems concerning branch-points and branch-cuts give additional information that will be useful in the discussion of special Riemann surfaces.

THEOREM I. *A free end of a branch-cut is a branch-point.*

Let α be a free end of a branch-cut C. Suppose that as z crosses this branch-cut it passes from the $k_1{}^{th}$ sheet into the $k_2{}^{th}$. Then as z makes a complete circuit about α starting from an initial position $z_0{}^{(k_1)}$ in the $k_1{}^{th}$ sheet it does not return at the end of the first revolution to that initial position, but it ends in a point $z_0{}^{(k_2)}$ in the $k_2{}^{th}$ sheet. Hence, from the definition of a branch-point, α is such a point.

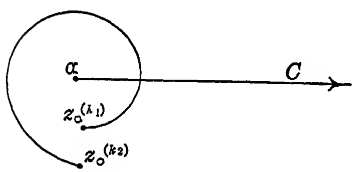

FIG. 116.

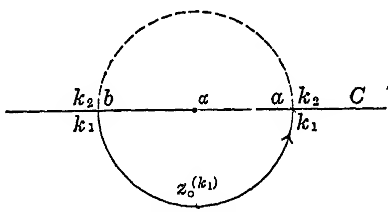

FIG. 117.

It is to be noted that in case a branch-cut ends in a point on the boundary of the region of existence it is not necessarily a branch-point.

THEOREM II. *If but one branch-cut passes through a branch-point, the connection of the sheets on the two sides of the branch-point is not the same.*

Let α be a branch-point through which but one branch-cut passes. The connection between the sheets can not be the same on the one side of α as on the other; for, in that case as the variable z describes a circuit about α it returns after each revolution to its initial position. For suppose the $k_1{}^{th}$ and $k_2{}^{th}$ sheets are connected along the given branch-cut C, as indicated in Fig. 117. If z has the initial

value $z_0^{(k_1)}$ and describes a circuit about α, say in the positive direction, then as z crosses C at a it passes from the k_1^{th} sheet to the k_2^{th}. Upon crossing the branch-cut again at b, z passes from the k_2^{th} sheet back to the k_1^{th} sheet returning to its initial value $z_0^{(k_1)}$.

THEOREM III. *If two branch-cuts, having different sequences of interchange of sheets associated with them, meet in a point, that point is either a branch-point or is an extremity of at least one other branch-cut.*

Let α be the common point of the two branch-cuts, I, II, Fig. 118. Along the branch-cut I, let the k_1^{th} sheet be associated with the k_2^{th} and along II, since the sequence of interchange of sheets can not be the same, suppose the k_2^{th} to be connected with a third sheet k_3^{th}. If the variable starts from an initial point $z_0^{(k_1)}$ in the k_1^{th} sheet, then by passing around α in a positive direction it passes into the k_2^{th} sheet, upon crossing I. Upon crossing II it passes from the k_2^{th} sheet to the k_3^{th} sheet and, if it crosses no further branch-cut, then instead of returning to the initial position $z_0^{(k_1)}$ at the close of the first circuit it ends in a point $z_0^{(k_3)}$ in the k_3^{th} sheet and α is a branch-point. Hence either α is a branch-point or there must be at least one more branch-cut like III ending in α by which the k_3^{th} sheet is connected with the k_1^{th} sheet.

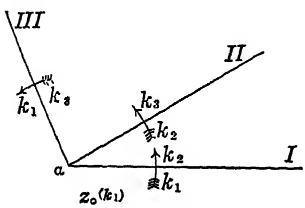

FIG. 118.

THEOREM IV. *If a change of sequence in the branches of a function occurs at any point of a branch-cut, then that point is a branch-point or it lies also on some other branch-cut.*

Since by hypothesis a change of sequence occurs at some point, say α, then as z describes a circuit about α it does not return to the sheet from which it started but passes into another sheet of the surface. Hence, the point α must either be a branch-point or another branch-cut must terminate at α.

THEOREM V. *A branch-cut passing through but one branch-point can not be a closed curve.*

If a branch-cut having but one branch-point is a closed curve, then by encircling that point along an arbitrarily small circuit, the variable returns to the same sheet; for, the connection between the sheets must be the same on both sides of the branch-point, since the portions of

the cut on the two sides of the point belong to the same branch-cut. It is, however, impossible for the variable to encircle a branch-point and not change sheets. Hence under the conditions set forth in the theorem, the branch-cut can not be a closed curve.

In general we have considered only paths which encircle a single branch-point. In Art. 60 we considered certain paths encircling two branch-points. We shall now consider the general effect of a path encircling two or more branch-points. We have the following theorem.

Theorem VI. *The effect of describing a closed circuit about several branch-points is the same as though the variable point had described a closed path about each of the branch-points in succession.**

We shall prove the theorem for the case of a circuit about three branch-points. The same argument holds for any finite number of such points. It is at once evident that a path can be deformed in any manner without affecting the result, provided in such a deformation a branch-point is not encountered. For by such a deformation no additional branch-cuts need be crossed an odd number of times. If crossed an even number of times, the final position of the variable point is in the same sheet as the initial point.

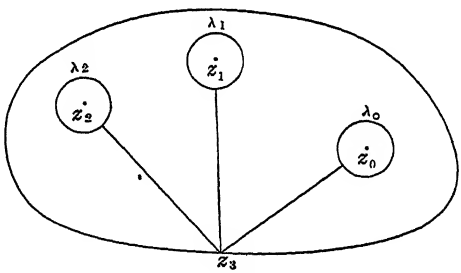

Fig. 119.

Consequently the closed path C about the three points z_0, z_1, z_2 can be deformed without affecting the final result into the succession of paths λ_0, λ_1, λ_2 about the three points z_0, z_1, z_2, respectively. Each of these paths begins and ends at z_3, as shown in Fig. 119, and consists of a small circle about the branch-point and a path connecting that circle with z_3. This connecting path, however, is traversed twice, once in each direction, so that any branch-cut crossed in going in one direction will be crossed again in an opposite direction when the path is traversed in the opposite direction. The effect of this por-

* The group of permutations which the function values undergo as the independent variable describes a closed path is often called a monodromic group. *Cf. Encyklopädie der Math. Wiss.*, Bd. II_2, p. 121.

tion of the path crossing a branch-cut can therefore be neglected. The total effect of traversing the closed path C is then the same as traversing in succession the closed circuits about the separate branch-points, as stated in the theorem.

The foregoing theorem also gives us a convenient way for testing the point $z = \infty$ for a branch-point. Consider a closed path inclosing all of the finite branch-points. If this path be traversed in a counter-clockwise direction the result is easily obtained by the theorem. However, since no finite branch-point lies exterior to this path, it follows that traversing the path in a clockwise direction is equivalent to encircling the point at infinity. Traversing the path in a clockwise direction gives the same interchange of sheets but in opposite order as traversing it in the opposite direction. Therefore, if the sheets are interchanged by traversing such a path in either direction the point $z = \infty$ is a branch-point.

Ex. 2. Given the function $w = \sqrt[3]{(z - z_0)(z - z_1)}$. Determine whether the point $z = \infty$ is a branch-point.

The given function has a branch-point of the second order at $z = z_0$ and at $z = z_1$. At z_0 and z_1 the branches interchange by successive clockwise revolutions about the point as follows

Before z changes,	$w_1,\ w_2,\ w_3$;
after one revolution,	$w_2,\ w_3,\ w_1$;
after two revolutions,	$w_3,\ w_1,\ w_2$;
after three revolutions,	$w_1,\ w_2,\ w_3$.

As z describes clockwise a closed circuit inclosing both z_0, z_1, we have:

Before z changes,	$w_1,\ w_2,\ w_3$;
after one revolution,	$w_3,\ w_1,\ w_2$;
after two revolutions,	$w_2,\ w_3,\ w_1$;
after three revolutions,	$w_1,\ w_2,\ w_3$.

The point $z = \infty$ is then a branch-point of order two.

It is to be observed that had the cyclic arrangement of the sheets at z_0, z_1 been such that these points might have been connected by a branch-cut, then there would have been no interchange of branches as z described a closed circuit about the two branch-points and consequently the point $z = \infty$ would not have been a branch-point. In order that the two points z_0 and z_1 might have been so connected the cyclic arrangement of the branches at this point would necessarily have involved the same sheets taken in opposite order. Such would be the case, for example, with the function

$$w = \sqrt[3]{\frac{z - z_0}{z - z_1}}.$$

64. Stereographic projection of a Riemann surface. As with single-valued functions, stereographic projection upon a sphere may

often be employed with advantage in the discussion of multiple-valued functions. The multiple-sheeted Riemann surface projects into a multiple-sheeted Riemann sphere whose sheets are associated at the branch-points. The branch-cuts in the plane go over into curves upon the sphere along which the variable point passes from one sheet to another.

In the case of the inverse of the exponential and trigonometric functions, namely the logarithmic and inverse trigonometric functions, the projection of the W-plane upon the sphere is also of interest. The infinite number of strips congruent with the fundamental strip map into similar regions having a common point and bounded by curves having a common tangent at the north pole. This result exhibits the fact that the exponential function e^w, in terms of which the other functions are defined, is a function which takes in the neighborhood of the essential singular point $w = \infty$ every complex value except zero and infinity. The branch-points may be regarded as boundary points of the region of existence on the sphere just as they are regarded upon the Riemann surface.

65. General properties of Riemann surfaces. Thus far we have considered only special cases of Riemann surfaces. In general to construct such a surface for a given function, we may proceed as follows.

1. *Determine the number of branches.* This number is equal to the largest number of distinct values which the function has for each value of the independent variable. In case the function is algebraic, the number of branches is equal to the degree of the algebraic equation defining the function, when that equation is freed from radicals.

2. *Locate the branch-points.* If the inverse of the given function is single-valued, the branch-points may be found by the theorem of Art. 59. In any case, they are to be found among the points of the complex plane representing values of the independent variable for which two or more values of the function are equal, and they may be either all at finite points or one of them may be the point at infinity. From among these points those that are branch-points may be found by permitting the independent variable to describe a closed path about each point and observing which of these paths leads to the initial value of the function after each circuit.

3. *Find the connection of the branches of the function at the various branch-points.* Having determined the number of branches and located the branch-points, the branches themselves are not as yet uniquely determined. This, however, is not necessary in order

that we may determine the connection of the branches. To show this connection find the cyclic permutation of the branches at each branch-point as the independent variable describes a closed path in the ordinary complex plane about that point. The number of branches affected at any branch-point is one greater than the order of the branch-point.

4. *Draw the branch-cuts.* The branch-cuts may be inserted in a variety of ways. They should be so chosen as suits best the purposes of the discussion in hand. When the cyclic permutation of the branches permits, the branch-cuts may connect the various branch-points, or when more convenient they may be drawn from the various branch-points to the point at infinity. As we have seen, a branch-cut need not be a straight line and may in fact be any ordinary curve that does not intersect itself. When once the branch-cuts are drawn, the various branches of the function are definitely determined aggregates of value-pairs (w, z), and with each sheet of the Riemann surface there is associated a definite branch of the function. The various sheets of the surface should be so connected along the branch-cuts that as the variable passes over one of these cuts the proper branches of the function interchange.

While the branches of a function are single-valued, it may happen that both w and z are multiple-valued functions of the other. In that case it is often convenient to introduce a third variable τ so related to w and z that the Riemann surfaces constituting the W-plane and the Z-plane, respectively, map continuously upon the single-sheeted τ-plane. Each branch of the given function $w = f(z)$ associated with a sheet of the Z-plane maps into a fundamental region of the τ-plane, which we may designate as a (z, τ) fundamental region. Likewise each sheet of the Riemann surface constituting the W-plane is associated with a branch of the inverse function $z = \phi(w)$ and maps upon a fundamental region of the τ-plane, which we may designate as a (w, τ) fundamental region. The (z, τ) regions do not coincide with the (w, τ) regions. As a result, that portion of the Riemann surface constituting the W-plane which corresponds to a sheet of the Z-plane, and hence to a branch of the function $w = f(z)$, does not coincide exactly with the whole of one or more sheets of the W-Riemann surface. It may be less or it may be more than one sheet, depending upon the nature of the functional relation between w and z. We shall refer to that portion of the W-Riemann surface which corresponds to the whole of a sheet of the Z-plane as a

fundamental region on the Riemann surface. When the branch-cuts are inserted, the correspondence between the points of this region and the particular sheet of the Z-plane is definitely determined; that is, this branch of the function $w = f(z)$ is fully determined. Since to a single sheet of the Z-plane there may correspond more than one sheet of the W-plane, it follows that some of the w-values may be repeated, although the branches of the given function remain single-valued. An illustrative example will aid to make clear the foregoing discussion.

Ex. 1. Consider the function

$$w^2 = z^3.$$

We introduce the auxiliary variable τ by putting

$$z = \tau^2, \qquad w = \tau^3.$$

The two-sheeted Z-Riemann surface required for this function maps upon the whole of the τ-plane, Fig. 120. If we take the negative half of the axis of reals as the branch-cut, then the upper sheet maps into the half of the τ-plane to the right of the axis of imaginaries, while the second sheet maps into the half of the plane to the left of the same axis. On the other hand, the three-sheeted W-plane maps

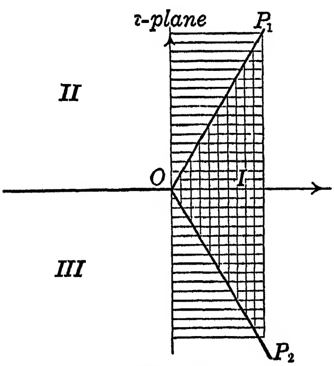

FIG. 120.

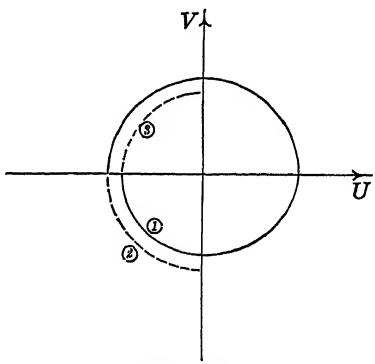

FIG. 121.

into the whole of the τ-plane as follows, where again we take the negative half of the axis of reals as the branch-cut. The first sheet maps into the region I bounded by the lines OP_1 and OP_2 making angles of $\frac{\pi}{3}$ and $-\frac{\pi}{3}$, respectively, with the positive axis of reals. The other two sheets map likewise into regions II and III, respectively. By direct comparison of the Z-surface and the W-surface, it will be seen that the region S of the W-surface corresponding to the first sheet of the Z-plane consists of the first sheet of the W-surface together with the second quadrant of the third sheet and the third quadrant of the second sheet, Fig. 121.

All of the values of S lying to the left of the axis of imaginaries are repeated, as will be seen from the figure; for, that portion of the region S lies in two sheets of the W-surface and the one directly over the other. However, no two of these values of w correspond to the same value of z; that is, while some of the values of the particular branch of the given function are repeated in S, nevertheless the branch is a single-valued function of z.

66. Singular points of multiple-valued functions. Since each branch of a multiple-valued analytic function is single-valued, if we exclude the branch-points from consideration we may, as we have already seen, regard such a function as an aggregate of single-valued functions, each of which may be holomorphic for those values of the independent variable which belong to the particular sheet of the Riemann surface with which the corresponding branch is associated. Aside from the branch-points, each branch of the given function may have such other singular points as any single-valued function. The singular points may affect one sheet or more than one sheet and consequently may be singular points of one or of more than one branch of the given function.

For example, a branch of a multiple-valued function may have a singular point which does not affect any other branch. As an illustration, consider the function

$$w = \log \log z. \tag{1}$$

Let $z = 1$. For this value of z, $\tau = \log z$ takes any one of the values

$$2k i\pi, \quad k = 0,\ 1,\ 2,\ \ldots.$$

These values of τ correspond to the value $z = 1$ in the various sheets of the Riemann surface constituting the Z-plane. For the sheet corresponding to $k = 0$, w becomes infinite for $z = 1$ and the function (1) ceases to be regular. For $k \neq 0$, however, the function can be expanded in powers of $(z - 1)$ and therefore $z = 1$ is a regular point for these sheets. Consequently, the given function has a singularity at $z = 1$, which, however, affects only one branch.

In the neighborhood of a point z_0 which is not a branch-point, a multiple-valued analytic function can be expanded in a series involving only integral powers of $(z - z_0)$. Such a point is a pole or an essential singular point according as the expansion contains a finite or an infinite number of terms having negative exponents. If there are no negative exponents, then z_0 is a regular point. Let us now examine the situation when z_0 is a branch-point. Suppose that at z_0 a finite number of branches, say k, of the given analytic function $w = f(z)$ are cyclically connected. Take a small region about z_0 bounded by a curve C closed upon the Riemann surface, such that it incloses no singular point nor branch-point other than z_0. Such

a curve must make k circuits about z_0 before it can be said to be closed. Let any convenient line extending out indefinitely from z_0 be taken as a branch-cut. We shall speak of the k-sheeted open region thus obtained on the Riemann surface, bounded by z_0 and C, as the region R. Denote the function defined in R by the k branches of $f(z)$ connected at z_0 by $F(z)$. By means of the substitution

$$z - z_0 = \tau^k,$$

the region R is mapped in a single-valued and continuous manner upon a region S of the one-sheeted τ-plane. Corresponding to the point z_0, we have the point $\tau = 0$, and except at the point z_0 itself, the mapping is conformal. The transformed function $\phi(\tau)$ corresponding to $F(z)$ is single-valued, and with at most the exception of the point $\tau = 0$ it is holomorphic in S. At $\tau = 0$, the function $\phi(\tau)$ may have a pole or an essential singularity. In either case, the limit of $\phi(\tau)$ as τ approaches zero does not exist. On the other hand the function $\phi(\tau)$ may approach a definite limiting value as τ approaches zero. If in the latter case we assign this limiting value as the value of $\phi(\tau)$ at $\tau = 0$, then, by virtue of Theorem I, Art. 51, the origin is a regular point of $\phi(\tau)$. In any case $\phi(\tau)$ can be expanded in the neighborhood of $\tau = 0$ by means of Laurent's expansion, thus obtaining

$$\phi(\tau) = \sum_{n=m}^{\infty} \alpha_n \tau^n, \tag{2}$$

which holds for values of τ exterior to an arbitrarily small circle about the origin and interior to any concentric circle within S. If $\phi(\tau)$ becomes infinite as τ approaches zero, then the origin is a pole and there are a finite number of negative terms in (2) equal in number to the order of the pole. If the pole is of order λ, then $m = -\lambda$. If on the other hand $\tau = 0$ is an essential singular point, m becomes $-\infty$. If it is a regular point, m is equal to or greater than zero and the expansion reduces to a Taylor series.

The character of $\phi(\tau)$ in the neighborhood of $\tau = 0$ enables us to determine the nature of $F(z)$, and hence of $f(z)$ for those branches affected at z_0. The character of those branches of $f(z)$ not connected at z_0 is quite independent of the existence of a branch-point at z_0 for the branches already considered. For those sheets not affected, this point may be a regular point or a pole or an essential singular point. It may also be a branch-point at which two or more of the remaining branches are associated with each other. The expansion of $F(z)$

in the deleted neighborhood of z_0 is obtained by replacing τ in (2) by $(z - z_0)^{\frac{1}{k}}$ giving

$$F(z) = \sum_{n=m}^{\infty} \alpha_n (z - z_0)^{\frac{n}{k}}, \tag{3}$$

which holds for values of z in R, that is in all of the sheets affected. As we know from Art. 8, there are k, k^{th} roots of $(z - z_0)$, namely:

$$(z - z_0)^{\frac{1}{k}}, \quad \omega(z - z_0)^{\frac{1}{k}}, \quad \omega^2(z - z_0)^{\frac{1}{k}}, \ \ldots, \quad \omega^{k-1}(z - z_0)^{\frac{1}{k}},$$

where $(z - z_0)^{\frac{1}{k}}$ is now restricted to the principal value of the root and ω is one of the imaginary k^{th} roots of unity. To get the expansion of the individual branches of $f(z)$ composing $F(z)$, all we need to do is to replace α_n in (3) by

$$\alpha_n, \ \alpha_n\omega^n, \ \ldots, \ \alpha_n(\omega^{k-1})^n,$$

respectively.

In case $\tau = 0$ is a regular point of $\phi(\tau)$, then from (2) we have

$$\underset{\tau \doteq 0}{L} \ \phi(\tau) = A,$$

where A is different from zero or equal to zero according as we have $m = 0$ or $m > 0$; hence, since z approaches z_0 as τ approaches zero, we have

$$\underset{z \doteq z_0}{L} \ F(z) = A.$$

Assigning this value as the value of $F(z)$ at z_0, then the branch-point z_0 is called a **point of continuity** of $F(z)$. The expansion (3) is in this case a series of increasing positive fractional powers of $(z - z_0)$, and we have $m \geqq 0$. If m is greater than zero, say equal to r, we say that $F(z)$ and hence $f(z)$ has a **zero point** of order r at z_0. The function w defined by $w^2 = z$ has a point of continuity at the branch-point $z = 0$. The expansion consists of one term, namely $z^{\frac{1}{2}}$, and the given function has a zero point of order one at this point.

If $\tau = 0$ is a pole of $\phi(\tau)$, m is negative and finite. Let the order of the pole at $\tau = 0$ be r. Then the expansion (2) and therefore (3) has r terms with negative exponents. We say that $F(z)$ has a **pole** of order r at z_0. The function w defined by the relation $w^2 = \frac{1}{z}$ furnishes an illustration. The point $z = 0$ is a branch-point, and at the same time it is a pole of order one, the expansion consisting of a single term having a negative exponent, namely $z^{-\frac{1}{2}}$.

If $\tau = 0$ is an essential singular point, then the expansion (2) contains an infinite number of terms with negative exponents; that is, we have $m = -\infty$. In this case, the transformed series (3) has also an infinite number of terms having negative fractional exponents. The point z_0 is called **an essential singular point** of $F(z)$. The function $w = e^{\frac{1}{\sqrt{z}}}$ has such a point at $z = 0$. This point is a simple branch-point. The form of the expansion in the deleted neighborhood of the origin is

$$w = 1 + z^{-\frac{1}{2}} + \frac{z^{-1}}{2!} + \frac{z^{-\frac{3}{2}}}{3!} + \cdots.$$

The origin is therefore an essential singular point as well as a branch-point.

Since the single-valued function $\phi(\tau)$ can not have other singular points than poles and essential singular points, it follows that the foregoing discussion exhausts the possibilities as regards the singularities of a function at a branch-point when a finite number of sheets are connected. That is, a branch-point may also be at the same time a point of continuity, a pole, or an essential singularity. In any case we shall class a branch-point among the singular points of a function, since the expansion of the function in the deleted neighborhood of the point involves fractional powers of the variable, that is, the function does not permit a proper power series development. Consequently, a branch-point is not to be included in the region of existence of the given function, and, moreover, we can not reach such a point by the process of analytic continuation from any regular point in the Riemann surface.

If an infinite number of sheets are connected at a branch-point, then the singularity at that point may be of a transcendental character.* For example, in the case of the logarithmic function, the origin is a singular point where log z becomes infinitely great by every possible approach of z to the origin. It is not, however, a pole because the order of the singularity is not finite. It is in this case sometimes spoken of as a logarithmic singularity. It is not within the scope of this volume to discuss the character of the various transcendental singularities that may occur at branch-points, further than to point out the illustration already cited.

We have excluded from the region of existence of an analytic

* Cf. Zoretti, *Leçons sur le prolongement analytique*, p. 61.

function the poles and essential singular points in the case of single-valued functions, and in the case of multiple-valued functions we have excluded also branch-points. It is often convenient to consider the region of existence as thus used, together with those singular points where w and z have a definite one-to-one correspondence. Thus in single-valued functions we may include in our consideration the poles of a function, if we associate with the pole $z = z_0$ the functional value $w = \infty$. Likewise in multiple-valued functions, we may include those branch-points at which the function has a point of continuity or a pole, by associating with the branch-point $z = z_0$ as the corresponding functional values, $w_0 = \underset{z=z_0}{L} f(z)$ and $w = \infty$, respectively. When these points of the Riemann surface and their corresponding functional values are added to the region of existence of an analytic function, we shall speak of the resulting aggregate of pair-values (w, z) of the given function as an **analytic configuration.** The essential singular points are not included in the notion of an analytic configuration; for, corresponding to such a singular point no particular w-point can be associated.

Just as in the consideration of single-valued analytic functions, the singular points are boundary points of the region of existence. In case of multiple-valued analytic functions, these singular points include the branch-points as well as the poles and essential singular points. When these boundary points constitute a closed curve upon the Riemann surface, then the region of existence of the given function has a **natural boundary;** that is, a boundary beyond which it is impossible to proceed by analytic continuations. The region of existence may be different in different sheets of the Riemann surface for the function.

67. Functions defined on a Riemann surface. Physical applications. The chief advantage of a Riemann surface is that it enables us not only to establish a one-to-one correspondence between the points of the Z-plane and those of the W-plane, but also to state that any continuous path in the one plane corresponds to a continuous path in the other. Regarding the branch-points as boundary points of the region of existence on the Riemann surface and using the term region as in the case of single-valued functions to mean an aggregate of inner points, unless otherwise stated, we can by means of Riemann surfaces extend to multiple-valued functions the general properties already discussed for single-valued functions. When use is made of Riemann surfaces, the mapping by means of multiple-

valued analytic functions is conformal in any region which contains no branch-points or other singular points of the function, provided that the derivative of the function is different from zero; that is, in any region in which the branches of the function are holomorphic and their derivatives do not vanish.

In the neighborhood of a branch-point z_0 of finite order, the expansion of a multiple-valued function requires an infinite series whose terms involve fractional powers of $(z - z_0)$. In the neighborhood of any isolated singular point other than a branch-point, the expansion of any branch takes the form of a Laurent series. In the neighborhood of a regular point, the branches of the function can each be expanded in a Taylor series. The circle of convergence may, however, lie partly in one sheet and partly in another depending upon the position of the point in whose neighborhood the function is expanded. In no case can the branch-point lie within the region of convergence; for, otherwise there would exist a regular power series development in the neighborhood of a branch-point, which we have seen is impossible. As a consequence it will be seen that by the process of analytic continuation a branch-point can not be included in the region of existence of an analytic function. It is for this reason, as already stated, that we have regarded branch-points as belonging to the boundary of the region of existence. Having thus excluded the branch-points, analytic continuation upon the Riemann surface can take place along any ordinary curve just as in the case of single-valued functions. The path of analytic continuation may lie wholly in one sheet or may pass from one sheet to another as the conditions require. Along any closed path of analytic continuation the terminal value of the function is identical with the initial value. If the path incloses a branch-point, then it must make as many circuits around that point as there are sheets connected at the point before the path can be said to be closed.

Similarly the process of integration can be extended to multiple-valued functions, the path of integration being any continuous ordinary curve upon the Riemann surface. In case the path does not inclose a branch-point, there is nothing new in the process. If, however, the path incloses a branch-point, say of order k, then it must pass $k+1$ times around that point before the path is closed. For example, suppose it to be required to integrate the function $w = \sqrt{z}$ along a closed path C about the origin. The closed path C

can be deformed into a double circle about the origin without affecting the value of the integral. We have then

$$\int_C z^{\frac{1}{2}}\,dz = \int_0^{\pi} \rho^{\frac{1}{2}} e^{i\frac{\theta}{2}} \cdot i\rho e^{i\theta} d\theta = i\rho^{\frac{3}{2}} \int_0^{4\pi} e^{i\frac{3\theta}{2}}\,d\theta$$

$$= i\rho^{\frac{3}{2}} \int_0^{4\pi} \cos\frac{3\,\theta}{2}\,d\theta - \rho^{\frac{3}{2}} \int_0^{4\pi} \sin\frac{3\,\theta}{2}\,d\theta = 0.$$

The **residue** of a multiple-valued function at an isolated singular point is defined, as in the case of single-valued functions, to be

$$\frac{1}{2\,\pi i}\int_C f(z)\,dz,$$

where C is a closed curve about the given singular point and inclosing no other singular point. In case the point z_0 is a branch-point at which k sheets are connected, the form of the expansion of the function is

$$f(z) = \sum_{n=m}^{\infty} \alpha_n (z - z_0)^{\frac{n}{k}},$$

where $m = -\lambda$ or $m = -\infty$ according as z_0 is a pole or an essential singular point. Since this series converges uniformly, it can be integrated term by term, and since C can be taken as a circle about z_0, traversed k times, we have as the residue at z_0

$$\frac{1}{2\,\pi i}\int_C f(z)\,dz = \frac{1}{2\,\pi i}\sum_{n=m}^{\infty} \alpha_n \int_0^{2\,k\pi} (z - z_0)^{\frac{n}{k}}\,dz$$

$$= k\alpha_{-k},$$

where $m \leqq -k$. When we have $-k < m < 0$, the residue is zero.

To find the residue when there is an isolated singular point at $z = \infty$, we have as the expansion of the function

$$f(z) = \sum_{n=m}^{\infty} \frac{\alpha_n}{z^{\frac{n}{k}}}, \qquad m < 0,$$

and hence obtain

$$\frac{1}{2\,\pi i}\int_C f(z)\,dz = -k\alpha_k.$$

It is often necessary to employ multiple-valued functions in mapping from one complex plane to another. In case the inverse function is single-valued, we can do as was done in Chapter IV, namely: we can map the whole of one plane upon a portion of the other plane, or we can now introduce a Riemann surface in place of

the one plane, thus making it possible to map in a continuous and single-valued manner the whole of either plane upon the whole of the other.

As an illustration, consider the function $w = z^2$. The W-plane is a double-sheeted Riemann surface having branch-points at $w = 0$ and $w = \infty$. Let us choose the negative half of the axis of reals as the branch-cut. By comparing Figs. 122 and 123, it will be seen that

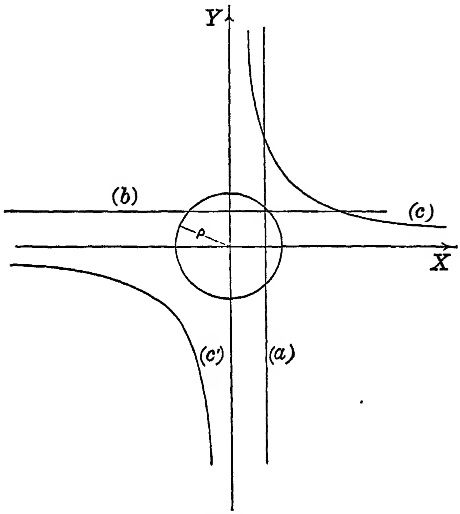

FIG. 122.

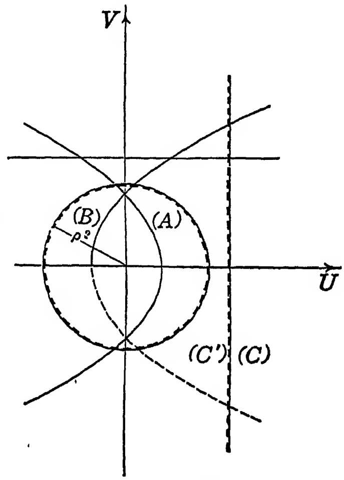

FIG. 123.

the half of the Z-plane lying to the right of the axis of imaginaries maps into the first sheet of the W-plane while the half of that plane to the left of this axis maps into the lower sheet of the Riemann surface constituting the W-plane.

Any line (a) lying in the right-hand half of the Z-plane and parallel to the Y-axis maps into the parabola (A) lying wholly in the first sheet of the Riemann surface. Likewise any line parallel to the Y-axis and lying in the left-hand half plane maps into a parabola lying wholly in the second sheet. The line (b) parallel to the X-axis maps into a parabola (B) situated symmetrically with respect to the U-axis and lying partly in the first sheet and partly in the second sheet as indicated, the dotted portion of the curve indicating that part which lies in the second sheet. A circle about the origin in the Z-plane maps into a double circle in the W-plane, one in each sheet. The

two circles thus obtained constitute a closed path as shown in Fig. 123. The line (C) situated in the first sheet and parallel to the V-axis maps into the branch (c) of an equilateral hyperbola lying in the first quadrant of the Z-plane. Similarly, the straight line (C') lying in the second sheet directly underneath (C) maps into the branch (c') of the same hyperbola lying in the third quadrant. Lines parallel to the U-axis map into hyperbolas cutting the first hyperbola at right angles.

It is of interest in this connection to point out some of the uses of Riemann surfaces in the discussion of physical problems. A solution of Laplace's differential equation gives a potential function. The solution may give a multiple-valued function u and the corresponding analytic function

$$w = u + iv = f(z)$$

to which it gives rise may have singular points other than the branch-points. In the corresponding physical problem, however, the potential must be single-valued when the boundary conditions are given.*

Similarly, if we arrive at the potential function through a process of integration, the function may be a cyclic function; that is, it may be multiple-valued because the path of integration encircles a singular point of the region under consideration. If such points be excluded from the region by arbitrarily small circles being drawn about them, the resulting region is multiply connected. It may be made simply connected by the introduction of barriers or cross-cuts, in which case the potential function obtained in this manner is single-valued. If the given region lies upon a Riemann surface, it will be seen that in order that the potential shall be single-valued in that region it is necessary to restrict the region to one sheet of the Riemann surface by excluding from the region all branch-points and branch-cuts.

Thus in Fig. 123, the transformation $w = z^2$ enables us to consider the potential in an electrostatic field bounded by the parabola A and lying exterior to this curve. This region corresponds to the region in the Z-plane lying to the right of the line a parallel to the Y-axis, Fig. 122. This transformation does not, however, enable us to consider a field upon the Riemann surface lying interior to the parabola, since this region contains the branch-point $w = 0$ and the negative U-axis. which we have here taken as a branch-cut.

*See Jeans, *Electricity and Magnetism*, Art. 330, also Lamb, *Hydrodynamics*, 3d Ed., Art. 62.

As another illustration, consider the function

$$w = \log \frac{z-1}{z+1}.$$

We have already discussed the mapping of this function where the amplitude θ of $\tau \equiv \frac{z-1}{z+1}$ is restricted to its chief amplitude. If we now allow θ to take all values we have a multiple-valued function and the corresponding Z-plane must consist of an infinite number of sheets. The branch-points of the function are located at $z = 1$, $z = -1$. All of the sheets of the surface are connected at these two points, but the cyclic arrangement is in the one case the reverse of the other. The two points can then be connected by a branch-cut. Having excluded the branch-points by arbitrarily small circles extending in each sheet in succession, the resulting surface is triply connected. If we now take the branch-cut as a barrier or cross-cut and draw an additional cross-cut from some point on this inner boundary of the surface to the point at infinity, we shall make the region simply-connected. Any region in any sheet exterior to a curve inclosing both of the two branch-points may therefore be taken as a suitable region for the consideration of a potential.

68. Function of a function. We have the following theorem.

THEOREM. *An analytic function of an analytic function is an analytic function.*

We shall demonstrate this theorem for the case of single-valued functions. From the discussion in this chapter it follows that the theorem may be extended to multiple-valued analytic functions.

Suppose $F(w)$ to be a single-valued analytic function of w, and let $w = f(z)$ likewise be a single-valued analytic function of z. We are to show that $F\{f(z)\} \equiv \phi(z)$ is also an analytic function of z. Let z_0 be any regular point of $f(z)$ such that $w_0 = f(z_0)$ is a regular point of the function $F(w)$. Then $F(w)$ can be expanded as a power series in $(w - w_0)$; that is, we may write

$$F(w) = \alpha_0 + \alpha_1(w - w_0) + \alpha_2(w - w_0)^2 + \cdots + \alpha_k(w - w_0)^k + \cdots. \quad (1)$$

Let the circle of convergence of (1) have a radius equal to ρ. Since $w = f(z)$ is analytic, it follows that $\alpha_k(w - w_0)^k$ is also analytic for all finite values of k and hence may be expanded as a power series in $(z - z_0)$. The circle of convergence of these power series may be denoted by r, since they have the same radius of convergence for all values of k. We may write

$$\begin{aligned}
\alpha_0 &= \beta_{0,0} \\
\alpha_1(w - w_0) &= \beta_{1,0} + \beta_{1,1}(z - z_0) + \beta_{1,2}(z - z_0)^2 + \cdots, \\
\alpha_2(w - w_0)^2 &= \beta_{2,0} + \beta_{2,1}(z - z_0) + \beta_{2,2}(z - z_0)^2 + \cdots, \\
\alpha_3(w - w_0)^3 &= \beta_{3,0} + \beta_{3,1}(z - z_0) + \beta_{3,2}(z - z_0)^2 + \cdots, \\
&\cdots\cdots\cdots\cdots\cdots,
\end{aligned}$$

Substituting these values in (1), we have the double series

$$\begin{aligned}\phi(z) &= \beta_{0,0}\\ &\quad + \beta_{1,0} + \beta_{1,1}(z-z_0) + \beta_{1,2}(z-z_0)^2 + \cdots\\ &\quad + \beta_{2,0} + \beta_{2,1}(z-z_0) + \beta_{2,2}(z-z_0)^2 + \cdots\\ &\quad + \beta_{3,0} + \beta_{3,1}(z-z_0) + \beta_{3,2}(z-z_0)^2 + \cdots . \qquad (2)\\ &\quad \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\end{aligned}$$

The rows of this series converge absolutely for all of those values of z for which $|z - z_0| = r' < r$, since each row is a power series with the radius of convergence r. Summing by rows, the resulting series each term of which is the sum of a row in (2) is none other than (1), which, as we know, also converges absolutely for all values of w for which $|w - w_0| = R < \rho$. Since the series formed by taking the absolute values of the terms of (2) converges by rows, it follows that the double series of these absolute values also converges *; that is, the given series converges absolutely. Consequently, by the theorem of Art. 44, we may sum it by columns as well as by rows. We have then

$$\phi(z) = \sum_0^\infty \beta_{k,0} + \sum_1^\infty \beta_{k,1}(z-z_0) + \sum_1^\infty \beta_{k,2}(z-z_0)^2 + \cdots + \sum_1^\infty \beta_{k,n}(z-z_0)^n + \cdots,$$

where we may put

$$\sum_0^\infty \beta_{k,0} = \beta_0, \qquad \sum_1^\infty \beta_{k,n} = \beta_n, \qquad n = 1, 2, 3, \ldots .$$

The function $\phi(z)$ is therefore represented by a power series in the neighborhood of z_0; hence, z_0 is a regular point of $\phi(z)$. But z_0 is any regular point of $w = f(z)$ for which the corresponding point w_0 is a regular point of $F(w)$. Hence, within the region for which z lies in the region of existence of $F(w)$, we may regard $\phi(z)$ as identical with $F\{f(z)\}$. It is possible, however, that the region of existence of the analytic function $\phi(z)$ thus defined may extend beyond the region of existence of $w = f(z)$. On the other hand, it is possible that the values of w given by the relation $w = f(z)$ may not lie within the region of existence of $F(w)$, in which case $F\{f(z)\}$ has no meaning.

69. Algebraic functions. In the preceding chapter, we discussed a special kind of algebraic functions, namely rational functions. In the present chapter we have had occasion to consider several particular algebraic functions. We shall now consider the

* See Bromwich, *Theory of Infinite Series*, Art. 31.

general case where $w = f(z)$ is defined by an irreducible equation of the form

$$F(w, z) \equiv w^n + f_1(z)\, w^{n-1} + f_2(z)\, w^{n-2} + \cdots + f_n(z) = 0,\ n > 0, \quad (1)$$

where $f_1(z), f_2(z), \ldots, f_n(z)$ are rational functions.* In some discussions it is convenient to write the foregoing equation in the following form

$$p_0(z)\, w^n + p_1(z)\, w^{n-1} + \cdots + p_k(z)\, w^{n-k} + \cdots + p_n(z) = 0, \quad (2)$$

where $p_k(z)$, $k = 0, 1, 2, \ldots, n$, is a rational integral function of z. For each value of z these equations have n roots and there are then in general n distinct values of w. We shall denote these values by $w_1, w_2, \ldots, w_n$. The function $w = f(z)$ thus defined is then a multiple-valued function, and $w_1, w_2, \ldots, w_n$ are all functions of z. In fact the functions $w_1, w_2, \ldots, w_n$ become the n branches of the given function when once the branch-cuts are properly chosen.

THEOREM I. *Every value of z_0 for which all of the n branches of an algebraic function remain finite and distinct is a regular point of each branch of the function.*

Corresponding to a circle C about z_0 as a center there may be drawn a circle C_k in the W-plane about each of the distinct points $w_{0,k}$ $(k = 1, 2, \ldots, n)$ as a center such that for all values of z in the region S bounded by C each of these circles C_k shall inclose values of w belonging to one and only one branch of the given function. Consequently, each branch of the function is single-valued for values of z in S.

We shall now show that z_0 is a regular point of each branch of the given algebraic function. To do this, it is sufficient to show that each of the functions w_k $(k = 1, 2, \ldots, n)$ admits of a derivative for values of z in S. Denote by Δw_k the increment of w_k corresponding to the increment Δz of z. If the given function is defined by the equation

$$F(w, z) = 0,$$

we have

$$\begin{aligned} \Delta F &= F(w + \Delta w, z + \Delta z) - F(w, z) \\ &= \frac{F(w + \Delta w, z + \Delta z) - F(w, z + \Delta z)}{\Delta w} \Delta w \\ &\quad + \frac{F(w, z + \Delta z) - F(w, z)}{\Delta z} \Delta z = 0. \end{aligned}$$

* For a somewhat different definition of an algebraic function, see Forsyth, *Theory of Functions*, 2d Ed., Art. 95. Compare also *Encyklopädie d. Math. Wiss.*, Vol. II, B_2, Art. 1.

But since the derivatives $\frac{\partial F}{\partial w}, \frac{\partial F}{\partial z}$ both exist, it follows that for $w = w_k$ the foregoing equation can be written in the form

$$\left\{\frac{\partial F}{\partial w_k} + \epsilon_1\right\} \Delta w_k + \left\{\frac{\partial F}{\partial z} + \epsilon_2\right\} \Delta z = 0,$$

where ϵ_1, ϵ_2 approach zero with $\Delta w_k, \Delta z$, respectively. We have from the foregoing relation

$$\frac{\Delta w_k}{\Delta z} = -\frac{\frac{\partial F}{\partial z} + \epsilon_2}{\frac{\partial F}{\partial w_k} + \epsilon_1}.$$

As Δz approaches zero, Δw_k also approaches zero and hence we have in the limit the same law as holds for the differentiation of implicit functions of real variables, namely

$$\frac{dw_k}{dz} = -\frac{\frac{\partial F}{\partial z}}{\frac{\partial F}{\partial w_k}}. \tag{3}$$

The value of $\frac{\partial F}{\partial w_k}$ is different from zero for all values of z in S; for, otherwise $F(w, z) = 0$ would have a multiple root * for some value of z in S, which is contrary to the hypothesis. The value of $\frac{dw_k}{dz}$ is given by (3) for any w_k, $k = 1, 2, \ldots, n$. Consequently, the point z_0 is a regular point for each of the functions w_k $(k = 1, 2, \ldots, n)$.

THEOREM II. *The number of points at which two or more of the branches of an algebraic function may become equal or infinite is finite.*

The finite values of z for which two or more of the values $w_1, w_2, \ldots, w_n$ become equal are those values of z that cause the discriminant of $F(w, z) = 0$ to vanish. Consequently, forming the resultant † R of the two polynomials

$$F(w, z), \qquad \frac{\partial F(w, z)}{\partial w},$$

the desired values of z are the roots of the equation obtained by equating R to zero. There can be at most a finite number of roots of this equation.

The finite values of z for which two or more of the values $w_1, w_2,$

* See Forsyth, *Theory of Functions*, 2d Ed., Art. 94.

† See Bôcher, *Introduction to Higher Algebra*, Art. 86.

$\ldots, w_n$ become infinite are among those for which the coefficient $p_0(z)$ vanishes. Since $p_0(z)$ is the least common multiple of the denominators of $f_1(z), f_2(z), \ldots, f_n(z)$ and therefore of finite degree in z, there can be only a finite number of roots of the equation $p_0(z) = 0$.

The only remaining z-point at which two or more values of $w_1, w_2, \ldots, w_n$ can become equal or infinite is the point $z = \infty$. Consequently, the total number of points at which the branches of an algebraic function can be infinite is finite in number, and hence the theorem.

From Theorem I it follows that any one of the branches w_k can be expanded in a Taylor series

$$w_k = \alpha_{0,k} + \alpha_{1,k}(z - z_0) + \alpha_{2,k}(z - z_0)^2 + \cdots, \tag{4}$$

which holds at least for all values of z in the region bounded by the circle of convergence C. The expansions for the various branches are of course different and in general the radius of convergence is not the same for all branches. We now see that there are only a finite number of points at which the branches of an algebraic function may become infinite or two or more of them be finite and equal. Since all other points must be regular points, it follows that every branch of an algebraic function is holomorphic except at a finite number of points, and hence we have the following theorem.

THEOREM III. *Every algebraic function is analytic and has only a finite number of singular points.*

The expansion (4) of any branch w_k in the neighborhood of a point where that branch is finite and distinct defines an element of the function, and from this element the algebraic function is completely and uniquely determined. It is also of interest to note that it follows from Theorem I that the singularities of an algebraic function can occur only at points where two or more of the branches have the same finite value or where one or more of the branches become infinite.

We shall use the expression **infinity of a function,** or more briefly an infinity, to mean a singular point of a multiple-valued function at which the function becomes infinite by at least one approach of the independent variable to the critical point. Infinities include both poles and essential singular points but exclude branch-points, unless those points are at the same time poles or essential singular points. The order of an infinity may therefore be either finite or

infinite. By the following theorem we shall show that the infinities of an algebraic function are necessarily poles.

THEOREM IV. *The infinities of an algebraic function are singular points of the coefficients and conversely. Moreover, an algebraic function can have only polar infinities.*

Let the given function $w = f(z)$ be defined by the algebraic equation

$$F(w, z) \equiv w^n + f_1(z)w^{n-1} + \cdots + f_n(z) = 0. \tag{5}$$

We shall first show that if z_0 is an infinity of any branch w_k, it is at the same time a singular point of at least one of the coefficients $f_1, f_2, \ldots, f_n$ of (5) and conversely. We shall also show that the order of the singularity of this coefficient is equal to or greater than the order of the infinity of w_k. From (5) we have by aid of the relation between the roots and coefficients of an algebraic equation

$$\begin{aligned} f_1(z) &= -(w_1 + w_2 + \cdots + w_k + \cdots + w_n), \\ f_2(z) &= w_1w_2 + w_1w_3 + \cdots + w_kw_2 + w_kw_3 + \cdots + w_{n-1}w_n; \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f_n(z) &= (-1)^n w_1w_2w_3 \ldots w_k \ldots w_n. \end{aligned}$$

Let the branches of the given function, other than w_k, be determined by an equation of the form

$$w^{n-1} + \phi_1(z)\, w^{n-2} + \phi_2(z)\, w^{n-3} + \cdots + \phi_{n-2}(z)\, w + \phi_{n-1}(z) = 0. \tag{6}$$

Then we may write

$$\left.\begin{aligned} f_1(z) &= -w_k + \phi_1(z), \\ f_2(z) &= -w_k\phi_1(z) + \phi_2(z), \\ &\cdots\cdots\cdots\cdots \\ f_{n-1}(z) &= -w_k\phi_{n-2}(z) + \phi_{n-1}(z), \\ f_n(z) &= -w_k\phi_{n-1}(z). \end{aligned}\right\} \tag{7}$$

From the last of equations (7) it follows that z_0 must be a singular point of $f_n(z)$ unless $\phi_{n-1}(z)$ has a zero point of as high order as the infinity of w_k. The singularity of $f_n(z)$ may, however, be of higher order than the infinity of w_k in case z_0 is also an infinity of $\phi_{n-1}(z)$. It may be of lower order provided $\phi_{n-1}(z)$ has a zero point at z_0 of order less than the order of the infinity of w_k. Consequently, $f_n(z)$ has a singular point at z_0 of order equal to or higher than the infinity of w_k unless z_0 is at the same time a zero point of $\phi_{n-1}(z)$. In the latter case, it follows in a similar manner from the next to the last of these equations that $f_{n-1}(z)$ has a singularity at z_0 of order equal

to or higher than the infinity of w_k unless $\phi_{n-2}(z)$ has a zero point at z_0. Continuing in this manner, it follows that either some one of the coefficients $f_2(z), \ldots, f_n(z)$ has an infinity at z_0 of at least as high an order as that of w_k, or $\phi_1(z)$ has a zero point at z_0. In the latter case, it follows from the first equation that $f_1(z)$ has an infinity at z_0 of the same order as w_k. We can conclude that at least one of the coefficients $f_1(z), f_2(z), \ldots, f_n(z)$ has a singularity at z_0 of equal or higher order than the infinity of w_k. However, the coefficients $f_1(z), f_2(z), \ldots, f_n(z)$ are all rational functions of z and therefore the function itself can have at most polar infinities.

Conversely, if some one of the coefficients $f_1(z), f_2(z), \ldots, f_n(z)$ has a singular point at z_0, then it must be a pole. From (7) it follows that if z_0 is not a pole of w_k, it must be a pole of at least one of the coefficients $\phi_1(z), \phi_2(z), \ldots, \phi_{n-1}(z)$ of (6). Then by similar reasoning z_0 is a pole of some branch, say $w_{k'}$, determined by (6) or of at least one of the coefficients of the resulting equation after both of the branches w_k and $w_{k'}$ have been removed. Continuing in this manner, it follows that either z_0 is a pole of some one of the first $n-1$ branches or of the coefficient in the last equation and hence of the last of the branches.

The poles which appear as the singular points of an algebraic function may at the same time be branch-points, but the function can have no other singularities than poles and branch-points, and as we have seen, an algebraic function can have in all at most a finite number of singular points. If the branch-point is at the same time a pole, the expansion of the function in an infinite series for values of z in the neighborhood of that point involves a finite number of terms having negative fractional exponents. If the pole occurs at a point other than a branch-point the expansion of any branch of the function involves a finite number of negative integral exponents. In the neighborhood of a branch-point which is not a pole the expansion involves only positive fractional exponents. In the neighborhood of any other finite point the expansion of each branch is accomplished by an ordinary Taylor series. The expansion of the function in the neighborhood of $z = \infty$ is in terms of $\dfrac{1}{z}$, the character of the expansion depending upon the nature of the function at that point.

We shall now consider whether every function $f(z)$ having no singular points other than poles and branch-points must be an alge-

braic function. Before discussing that question, however, we must demonstrate the following proposition.

THEOREM V. *If in a given region S of the complex plane, a function $w = f(z)$ has no singular points other than poles and branch-points and for values of z in S has m branches, which are in general distinct, then any symmetric polynomial of these branches is in S a meromorphic function of z.*

Denote the m branches of w by $w_1, w_2, \ldots, w_m$. For any value of z these branches of w may remain distinct, or two or more of them may take the same finite value, or finally one or more of them may become infinite. These branches are m single-valued functions of z when we consider the branch-cuts as replaced by barriers restricting the variation of z. From the given hypothesis it follows that each w_k admits of a derivative except at certain singular points, and hence in that portion of the region S exclusive of singular points w_k is holomorphic.

The expansion of w_k in the neighborhood of any point z_0 which is not a pole or a branch-point is of the form

$$w_k = \alpha_0 + \alpha_1(z - z_0) + \alpha_2(z - z_0)^2 + \cdots + \alpha_n(z - z_0)^n + \cdots. \quad (8)$$

In case z_0 is a branch-point of the function lying in S, where r values of w become equal but remain finite, the expansion of the given function for those branches is obtained by introducing the auxiliary function

$$z - z_0 = \tau^r.$$

As we have seen, this substitution leads to an expansion of the given function in the neighborhood of z_0 of the form

$$\alpha_0 + \alpha_1(z - z_0)^{\frac{1}{r}} + \alpha_2(z - z_0)^{\frac{2}{r}} + \cdots + \alpha_n(z - z_0)^{\frac{n}{r}} + \cdots. \quad (9)$$

No terms with negative exponents appear in the expansion since the branch-point is not an infinity. To get the expansion in the neighborhood of z_0 of each of the r branches associated at that point, we need to regard $(z - z_0)^{\frac{n}{r}}$, $n = 1, 2, 3, \ldots$, as the principal value of the r^{th} root of $(z - z_0)^n$ and replace α_n by

$$\alpha_n, \ \alpha_n\omega^n, \ \alpha_n(\omega^2)^n, \ \ldots, \ \alpha_n(\omega^{r-1})^n,$$

where ω is one of the r imaginary r^{th} roots of unity, as explained in Art. 66. The remaining $m - r$ branches may remain distinct or form by themselves one or more cycles.

Any symmetric polynomial of the m branches $w_1, w_2, \ldots, w_m$ can be expressed in terms of the sums of equal powers of the w_k's; that is, in terms of functions of the type *

$$S_a = w_1{}^a + w_2{}^a + \cdots + w_m{}^a.$$

In adding equal powers of w_k's, however, the coefficients of terms having fractional exponents vanish by virtue of the relation

$$1 + \omega + \omega^2 + \cdots + \omega^{r-1} = 0.$$

Hence, in this case also, the expansion of any symmetric polynomial of $w_1, w_2, \ldots, w_m$ involves only positive integral powers of $(z - z_0)$, and consequently z_0 is a regular point of the given function $f(z)$.

Finally, let us suppose z_0 is a pole of the given function. In this point then one or more values of w_k become infinite as z approaches z_0. In the latter case the point z_0 may be a branch-point. The expansion of w_k in the neighborhood of such a point involves a finite number of terms with negative exponents, which may be either fractional or integral. In forming a symmetric function of $w_1, w_2, \ldots, w_m$ the exponents all become integral even in case z_0 is a branch-point and hence the resulting function has a pole at z_0.

As we now see, any symmetric function of $w_1, w_2, \ldots, w_m$ can have in the region S only polar singularities. There can be but a finite number of poles in S, for otherwise there must be at least one essential singular point. Consequently, any symmetric function of $w_1, w_2, \ldots, w_m$ must be meromorphic in the given region S as the theorem requires.

We shall now consider the following theorem.

THEOREM VI. *Every analytic function $w = f(z)$ having n values for each value of z and having in the entire complex plane no other singularities than poles and branch-points can be expressed as a root of an algebraic equation of degree n in w, the coefficients of which are rational functions of z, and consequently $w = f(z)$ is an algebraic function.*

Corresponding to any point z_0 of the complex plane, it follows from the hypothesis set forth in the theorem that w has n values, which as before we denote by $w_1, w_2, \ldots, w_n$. These values are in general distinct.

* See Bôcher, *Introduction to Higher Algebra*, p. 241.

We shall now consider the following symmetric polynomials of $w_1, w_2, \ldots, w_n$.

$$\begin{aligned} f_1(z) &= -(w_1 + w_2 + \cdots + w_n), \\ f_2(z) &= w_1w_2 + w_1w_3 + \cdots + w_{n-1}w_n, \\ &\cdots\cdots\cdots\cdots \\ f_n(z) &= (-1)^n w_1w_2 \ldots w_n. \end{aligned}$$

From Theorem V, it follows that the functions $f_1(z), f_2(z), \ldots, f_n(z)$ are meromorphic for all values of z, since the region S may now include the whole complex plane. These functions therefore have only polar singularities and must be rational functions.

To complete the proof, we have only to equate to zero the product $\prod_{k=1}^{n} (w - w_k)$, namely

$$(w - w_1)(w - w_2) \ldots (w - w_n) = 0. \tag{10}$$

By the principles of elementary algebra, this equation can be written in the form

$$w^n + f_1(z)\, w^{n-1} + f_2(z)\, w^{n-2} + \cdots + f_n(z) = 0.$$

The function $w = f(z)$ is a root of this equation, and as we have seen the coefficients $f_1(z), f_2(z), \ldots, f_n(z)$ are rational functions. Hence the theorem.

The Riemann surface for an algebraic function can now be determined. If the equation $F(w, z) = 0$ defining the function is of the degree n in w then the Z-plane must be an n-sheeted surface. The branch-points and the connection of the sheets at each can be determined by the methods given. The branch-cuts can be drawn between branch-points where the same sheets are affected but their cyclic arrangements are of opposite order, or they may be drawn to the point $z = \infty$. All of the sheets must form a connected surface since by definition $F(w, z) = 0$ is an irreducible equation. In case this equation is reducible, that is, can be separated into two or more factors involving w and z, then it defines not one algebraic function but two or more such functions and the corresponding Riemann surface separates into two or more distinct surfaces. In fact, a given relation between w and z may be said to represent one multiple-valued function or more than one function according as the corresponding Riemann surface consists of one connected part or of two distinct parts.

EXERCISES

1. Show that z^α, where $\alpha \equiv a + ib$, is a multiple-valued analytic function with a branch-point of an infinitely high order at the origin.

HINT: Put $z^\alpha = e^{\alpha \log z}$.

2. Given the analytic function $w = \dfrac{1}{\sqrt{2 + z^2}}$. Locate the branch-points and determine whether they are at the same time poles.

3. Evaluate the integral $\displaystyle\int_C \frac{dz}{\sqrt{z^2 + 1}}$, where C is a curve closed on the Riemann surface about the point $z = i$.

4. Knowing that

$$\arc \tan z = \int_0^z \frac{dz}{1 + z^2},$$

expand in an infinite series the function $f(z) = \arc \tan z$. How large is the circle of convergence of this series? What singular points restrict the size of the circle of convergence? Are these singular points also branch-points?

5. Determine the branch-points of the function $w = \arc \sin z$.

6. By aid of the definitions of circular functions, show that

$$w = \arc \tan z = \frac{1}{2i} \log \frac{1 + iz}{1 - iz}.$$

Locate the branch-points of w and determine a region on the Riemann surface in which each of the infinite number of branches is holomorphic. Show whether this region is simply or multiply connected.

7. Show that the function $f(z) = \log (\sin z)$ is analytic.

8. Discuss the Riemann surface for the function $w = \sqrt{(z - \alpha)(z - \beta)^2}$. What physical phenomena does this functional relation represent?

9. Construct a model showing the connection of the sheets of the Riemann surface required for the function discussed in Art. 61.

10. Discuss the Riemann surface for the function $w^2 - 1 = z^3$. Determine a fundamental region on the W-Riemann surface.

11. By mapping the Z-plane upon the W-plane by means of the function $w = e^z$, show that w takes every value, except zero and infinity, in the deleted neighborhood of $z = \infty$.

12. Given the algebraic function, $w^3 - 18w - 35z = 0$. Determine the character of the Riemann surface suited to this function by the method of Art. 60.

13. Expand the function $w = \sqrt{z}$ in an infinite series for values of z in the neighborhood of $z = \alpha$, where $\alpha = 2e^{i\frac{\pi}{3}}$.

14. Given the function $w = \sqrt{1 - z^3}$. Examine the character of this function in the neighborhood of $z = \infty$.

15. Indicate the form of the expansion in an infinite series which the analytic function $w = f(z)$ takes for values of z in the neighborhood of $z = i$, (*a*) if $z = i$ is a point of continuity and a branch-point of order 2; (*b*) if $z = i$ is a pole of order k but not a branch-point; (*c*) if $z = i$ is a branch-point where 3 branches come together and at the same time a pole of order 4; (*d*) if $z = i$ is a branch-point of order 6 and an essential singular point; (*e*) if $z = i$ is a zero-point of order 3 and a branch-point where 2 branches become equal.

16. Distinguish between an algebraic function and a transcendental function (*a*) as to the number of possible zero points, (*b*) as to the number of poles and essential singular points that may occur, (*c*) as to the number and order of the branch-points that the function may have. Illustrate in each case by a particular function.

17. Does the expression arc sin (cos z) represent a multiple-valued analytic function or a number of single-valued analytic functions?

18. Discuss the Riemann surfaces for the following functions:

$$(a)\quad w = \sqrt{z-i} + \sqrt{z-2},$$

$$(b)\quad w = \sqrt{z-1} + \frac{1}{\sqrt{z+2}},$$

$$(c)\quad w^4 - 4w = z.$$

INDEX

INDEX

(Numbers refer to pages.)